RENCANA PEMBELAJARAN SEMESTER (RPS)
PROGRAM STUDI : S2 MATEMATIKA
FAKULTAS / PPs: MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS ANDALAS

MATA KULIAH	KODE	RUMPUN MK	BOBOT	SEMESTER	TANGGAL PENYUSUNAN
Topologi Aljabar	**PAM 638**	**ALJABAR (PILIHAN)**	**3 SKS**	**TIGA**	**16 FEBRUARI 2016**

OTORISASI	**DOSEN PENGEMBANG RPS**	**KOORDINATOR RUMPUN MK**	**KETUA PROGRAM STUDI**
	Dr. Admi Nazra	Dr. Yanita	Dr. Muhafzan

Capaian Pembelajaran (CP)	**CPL Program Studi / Kompetensi**	
	KU1, KU2, KL1, KL2	
	CP Mata Kuliah	
	1	Memahami dan menguasai konsep Homeomorpisma dan Ekivalen Homotopi
	2	Memahami dan menguasai konsep Homologi
	3	Memahami dan menguasai konsep Cohomologi
	4	Memahami dan menguasai konsep Fiber bundle dan vector bundle.
Deskripsi Singkat MK	Kuliah ini merupakan suatu kajian dari ilmu topologi aljabar, yang disajikan melalui contoh-contoh sederhana yang dibangun pada sel-sel kompleks. Topik-topik yang disajikan dalam perkuliahan ini meliputi homeomorpisma, ekivalen homotopi, torus, bidang Mobius, permukaan tertutup, botol Klein, sel kompleks, grup fundamental, grup homotopi, grup homologi, grup cohomologi, fiber bundle, vektor bundle, barisan spektral dan kelas-kelas karakteristik. Dengan mengikuti kuliah ini, mahasiswa diharapkan mempunyai pengetahuan dasar untuk mendalami ilmu aljabar topologi bilamana dikemudian hari mereka melanjutkan studi di program doktor di bidang topologi.	
Bahan Kajian	Homeomorpisma dan Ekivalen Homotopi, Homologi, Cohomologi, Fiber bundle dan vector bundle	
Pustaka	**Utama :**	
	H. Sato, *Algebraic Topology: An Intuitive Approach*, Translations of Mathematical Monographs, AMS, Vol 183, 1999.	
	Pendukung :	
	V.A. Vassiliev, *Introduction to Topology*, Student Mathematical Library, AMS, Vol 14, 2000.	
Media Pembelajaran	**Perangkat lunak :**	**Perangkat keras :**

<table>
<tr><td></td><td>-</td><td>Komputer/Laptop dan LCD Projector</td></tr>
<tr><td>Team Teaching</td><td colspan="2">Dr. Admi Nazra</td></tr>
<tr><td>Assessment</td><td colspan="2">
<table>
<tr><th>NO</th><th>KOMPONEN PENILAIAN</th><th>BOBOT (%)</th></tr>
<tr><td colspan="3">Penilaian Hasil</td></tr>
<tr><td>1</td><td>Ujian Tengah Semester</td><td>30 %</td></tr>
<tr><td>2</td><td>Ujian Akhir Semester</td><td>30 %</td></tr>
<tr><td colspan="3">Penilaian Proses</td></tr>
<tr><td>1</td><td>Kemampuan berpikir kritis dan berargumen logis</td><td>20 %</td></tr>
<tr><td>2</td><td>Kemampuan analitis</td><td>10 %</td></tr>
<tr><td>3</td><td>Kerjasama dalam tim</td><td>10 %</td></tr>
<tr><td colspan="2">TOTAL</td><td>100 %</td></tr>
</table>
</td></tr>
<tr><td>Norma Akademik</td><td colspan="2">
1. Mengikuti Peraturan Akademik Program Pascasarjana Universitas Andalas.
2. Kegiatan pembelajaran sesuai jadwal yang telah ditetapkan, toleransi keterlambatan 15 menit.
3. Selama proses pembelajaran berlangsung, handphone dimatikan.
4. Pengumpulan tugas ditetapkan sesuai jadwal dan dilakukan sebelum pembelajaran dimulai. Jika tugas dikumpulkan:
• Setelah pembelajaran dimulai, maka penilaian dikurangi menjadi 75%
• Terlambat 1 (satu) hari, maka penilaian dikurangi menjadi 50%
• Terlambat lebih dari 1 (satu) hari, maka mahasiswa dianggap tidak mengerjakan tugas
5. Jika mahasiswa melakukan kecurangan, baik dalam absensi, tugas maupun ujian, maka mahasiswa tersebut dianggap tidak lulus.
6. Aturan jumlah minimal presensi dalam pembelajaran tetap diberlakukan, termasuk aturan cara berpakaian
</td></tr>
<tr><td>Matakuliah Prasyarat</td><td colspan="2">tidak ada</td></tr>
</table>

MINGGU	SUB-CP-MK	INDIKATOR	KRITERIA DAN BENTUK PENILAIAN	METODE PEMBELAJARAN	MATERI PEMBELAJARAN	KET
1	• Kemampuan memahami kapan dua ruang topologi dikatakan saling homeomorpik, dan kapan suatu fungsi dakatakan homotopi ekivalen.	• Kedisiplinan dalam melaksanakan kontrak kuliah • Ketepatan dalam memahami materi terkait	Keaktifan	Presentasi dan Diskusi	Homeomorpisma dan Ekivalen Homotopi	
2	• Kemampuan memahami jenis-jenis ruang topologi, dan bagaimana membentuk ruang topologi yang baru dari ruang yang sudah diketahui.	• Ketepatan dalam memahami materi terkait • Ketepatan dalam mengerjakan tugas • Kerapihan dan keaslian tugas	Keaktifan dan tugas rutin	Presentasi dan Diskusi	Ruang topologi dan sel kompleks	
3	• Kemampuan mengidentifiksai himpunan-himouna yang homotopik dan mampu menghitung grup fundamental dari suatu ruang topologi.	• Ketepatan dalam memahami materi terkait • Ketepatan dalam mengerjakan tugas • Kerapihan dan keaslian tugas	Keaktifan dan tugas rutin	Presentasi dan Diskusi	Homotopi	
4	• Kemampuan mengembangkan konsep homotopi di ruang berdimensi rendah ke ruang berdimensi tinggi.	• Ketepatan dalam memahami materi terkait	Keaktifan	Presentasi dan Diskusi	Homotopi	
5	Untuk mengevaluasi meteri yang sudah diperoleh	• Ketepatan dalam memahami materi terkait • Ketepatan dalam mengerjakan tugas	Keaktifan dan tugas rutin	Presentasi dan Diskusi	Ujian I	

6	• Kemampuan menghitung grup homologi dari suatu ruang topologi serta mengaplikasikan aksioma terkait.	• Ketepatan dalam memahami materi terkait • Ketepatan dalam mengerjakan tugas • Kerapihan dan keaslian tugas	Keaktifan dan tugas rutin	Presentasi dan Diskusi	Homologi	
7	• Kemampuan memahami akibat dari aksioma sebelumnya.	• Ketepatan dalam memahami materi terkait	Keaktifan	Presentasi dan Diskusi	Homologi	
8 - 9	**UJIAN TENGAH SEMESTER**					
10	• Kemampuan memahami definisi grup homologi dari sel kompleks dan simplisial kompleks.	• Ketepatan dalam memahami materi terkait • Ketepatan dalam mengerjakan tugas • Kerapihan dan keaslian tugas	Keaktifan dan tugas rutin	Presentasi dan Diskusi	Grup homologi dari sel kompleks	
11	• Kemampuan menghitung grup homologi dari sel kompleks.	• Ketepatan dalam memahami materi terkait • Ketepatan dalam mengerjakan tugas • Kerapihan dan keaslian tugas	Keaktifan dan tugas rutin	Presentasi dan Diskusi	Grup homologi dari sel kompleks	
12	• Kemampuan menghitung grup cohomologi dari ruang topologi.	• Ketepatan dalam memahami materi terkait	Keaktifan	Presentasi dan Diskusi	Cohomologi	
13	• Kemampuan menghitung grup cohomologi dari simplisial kompleks.	• Ketepatan dalam memahami materi terkait • Ketepatan dalam mengerjakan tugas	Keaktifan dan tugas rutin	Presentasi dan Diskusi	Cohomologi	

		• Kerapihan dan keaslian tugas				
14	• Kemampuan menghitung hasil kali dari grup abelian dan mengaplikasikan rumus Kunneth.	• Ketepatan dalam memahami materi terkait	Keaktifan dan tugas rutin	Presentasi dan Diskusi	Teorema Koefisien universal.	
15	• Kemampuan menjelaskan konsep Cup Product dan Teorema Koefisien universal.	• Ketepatan dalam memahami materi terkait • Ketepatan dalam mengerjakan tugas • Kerapihan dan keaslian tugas	Keaktifan dan tugas rutin	Presentasi dan Diskusi	Teorema Koefisien universal.	
16	• Kemampuan menjelaskan konsep Fiber bundle, Vektor bundle. Grassmann manifolds.	• Ketepatan dalam memahami dan menjelaskan kaitan antara topik-topik yang telah dikaji	Keaktifan dan tugas rutin	Presentasi dan Diskusi	Fiber bundle dan vector bundle.	
17	**UJIAN AKHIR SEMESTER**					

Introduction to Modern Topology and Geometry

Anatole Katok

Alexey Sossinsky

Contents

CHAPTER 1

BASIC TOPOLOGY

Topology, sometimes referred to as "the mathematics of continuity", or "rubber sheet geometry", or "the theory of abstract topological spaces", is all of these, but, above all, it is a *language*, used by mathematicians in practically all branches of our science. In this chapter, we will learn the basic words and expressions of this language as well as its "grammar", i.e. the most general notions, methods and basic results of topology. We will also start building the "library" of examples, both "nice and natural" such as manifolds or the Cantor set, other more complicated and even pathological. Those examples often possess other structures in addition to topology and this provides the key link between topology and other branches of geometry. They will serve as illustrations and the testing ground for the notions and methods developed in later chapters.

1.1. Topological spaces

The notion of topological space is defined by means of rather simple and abstract axioms. It is very useful as an "umbrella" concept which allows to use the geometric language and the geometric way of thinking in a broad variety of vastly different situations. Because of the simplicity and elasticity of this notion, very little can be said about topological spaces in full generality. And so, as we go along, we will impose additional restrictions on topological spaces, which will enable us to obtain meaningful but still quite general assertions, useful in many different situations in the most varied parts of mathematics.

1.1.1. Basic definitions and first examples.

DEFINITION 1.1.1. A *topological space* is a pair $(X, \mathcal{T})$ where X is a set and $\mathcal{T}$ is a family of subsets of X (called the *topology* of X) whose elements are called *open sets* such that

(1) $\varnothing, X \in \mathcal{T}$ (the empty set and X itself are open),
(2) if $\{O_\alpha\}_{\alpha \in A} \subset \mathcal{T}$ then $\bigcup_{\alpha\in A} O_\alpha \in \mathcal{T}$ for any set A (the union of any number of open sets is open),
(3) if $\{O_i\}_{i=1}^k \subset \mathcal{T}$, then $\bigcap_{i=1}^k O_i \in \mathcal{T}$ (the intersection of a finite number of open sets is open).

If $x \in X$, then an open set containing x is said to be an (*open*) *neighborhood* of x.

We will usually omit $\mathcal{T}$ in the notation and will simply speak about a "topological space X" assuming that the topology has been described.

The complements to the open sets $O \in \mathcal{T}$ are called *closed* sets .

EXAMPLE 1.1.2. Euclidean space $\mathbb{R}^n$ acquires the structure of a topological space if its open sets are defined as in the calculus or elementary real analysis course (i.e a set $A \subset \mathbb{R}^n$ is open if for every point $x \in A$ a certain ball centered in x is contained in A).

EXAMPLE 1.1.3. If all subsets of the integers $\mathbb{Z}$ are declared open, then $\mathbb{Z}$ is a topological space in the so–called discrete topology.

EXAMPLE 1.1.4. If in the set of real numbers $\mathbb{R}$ we declare open (besides the empty set and $\mathbb{R}$) all the half-lines $\{x \in \mathbb{R} | x \geq a\}$, $a \in \mathbb{R}$, then we do not obtain a topological space: the first and third axiom of topological spaces hold, but the second one does not (e.g. for the collection of all half lines with positive endpoints).

EXAMPLE 1.1.5. Example 1.1.2 can be extended to provide the broad class of topological spaces which covers most of the natural situations.

Namely, a *distance function* or a *metric* is a function of two variables on a set X (i,e, a function of the Cartesian product $X \times X$ of X with itself) which is nonnegative, symmetric, strictly positive outside the diagonal, and satisfies the triangle inequality (see Definition 3.1.1). Then one defines an (open) ball or radius $r > 0$ around a point $x \in X$ as the set of all points at a distance less that r from X, and an open subset of X as a set which together with any of its points contains some ball around that point. It follows easily from the properties of the distance function that this defines a topology which is usually called the *metric topology*. Naturally, different metrics may define the same topology. We postpone detailed discussion of these notions till Chapter 3 but will occasionally notice how natural metrics appear in various examples considered in the present chapter.

The *closure* $\bar{A}$ of a set $A \subset X$ is the smallest closed set containing A, that is, $\bar{A} := \bigcap\{C \mid A \subset C \text{ and } C \text{ closed}\}$. A set $A \subset X$ is called *dense* (or *everywhere dense*) if $\bar{A} = X$. A set $A \subset X$ is called *nowhere dense* if $X \setminus \bar{A}$ is everywhere dense.

A point x is said to be an *accumulation point* (or sometimes *limit point*) of $A \subset X$ if every neighborhood of x contains infinitely many points of A.

A point $x \in A$ is called an *interior point* of A if A contains an open neighborhood of x. The set of interior points of A is called the *interior* of A and is denoted by $\operatorname{Int} A$. Thus a set is open if and only if all of its points are interior points or, equivalently $A = \operatorname{Int} A$.

A point x is called a *boundary point* of A if it is neither an interior point of A nor an interior point of $X \setminus A$. The set of boundary points is called the *boundary* of A and is denoted by ∂A. Obviously $\bar{A} = A \cup \partial A$. Thus a set is closed if and only if it contains its boundary.

EXERCISE 1.1.1. Prove that for any set A in a topological space we have $\partial \overline{A} \subset \partial A$ and $\partial(\operatorname{Int} A) \subset \partial A$. Give an example when all these three sets are different.

A sequence $\{x_i\}_{i \in \mathbb{N}} \subset X$ is said to *converge* to $x \in X$ if for every open set O containing x there exists an $N \in \mathbb{N}$ such that $\{x_i\}_{i>N} \subset O$. Any such point x is called a *limit* of the sequence.

EXAMPLE 1.1.6. In the case of Euclidean space $\mathbb{R}^n$ with the standard topology, the above definitions (of neighborhood, closure, interior, convergence, accumulation point) coincide with the ones familiar from the calculus or elementary real analysis course.

EXAMPLE 1.1.7. For the real line $\mathbb{R}$ with the discrete topology (all sets are open), the above definitions have the following weird consequences: any set has neither accumulation nor boundary points, its closure (as well as its interior) is the set itself, the sequence $\{1/n\}$ does not converge to 0.

Let $(X, \mathcal{T})$ be a topological space. A set $D \subset X$ is called *dense* or *everywhere dense* in X if $\bar{D} = X$. A set $A \subset X$ is called *nowhere dense* if $X \setminus \bar{A}$ is everywhere dense.

The space X is said to be *separable* if it has a finite or countable dense subset. A point $x \in X$ is called *isolated* if the one–point set $\{x\}$ is open.

EXAMPLE 1.1.8. The real line $\mathbb{R}$ in the discrete topology is *not* separable (its only dense subset is $\mathbb{R}$ itself) and each of its points is isolated (i.e. is not an accumulation point), but $\mathbb{R}$ *is* separable in the standard topology (the rationals $\mathbb{Q} \subset \mathbb{R}$ are dense).

1.1.2. Base of a topology. In practice, it may be awkward to list *all* the open sets constituting a topology; fortunately, one can often define the topology by describing a much smaller collection, which in a sense generates the entire topology.

DEFINITION 1.1.9. A *base* for the topology $\mathcal{T}$ is a subcollection $\beta \subset \mathcal{T}$ such that for any $O \in \mathcal{T}$ there is a $B \in \beta$ for which we have $x \in B \subset O$.

Most topological spaces considered in analysis and geometry (but not in algebraic geometry) have a *countable base*. Such topological spaces are often called *second countable*.

A *base of neighborhoods of a point* x is a collection $\mathcal{B}$ of open neighborhoods of x such that any neighborhood of x contains an element of $\mathcal{B}$.

If any point of a topological space has a countable base of neighborhoods, then the space (or the topology) is called *first countable*.

EXAMPLE 1.1.10. Euclidean space $\mathbb{R}^n$ with the standard topology (the usual open and closed sets) has bases consisting of all open balls, open balls of rational radius, open balls of rational center and radius. The latter is a countable base.

EXAMPLE 1.1.11. The real line (or any uncountable set) in the discrete topology (all sets are open) is an example of a first countable but not second countable topological space.

PROPOSITION 1.1.12. *Every topological space with a countable space is separable.*

PROOF. Pick a point in each element of a countable base. The resulting set is at most countable. It is dense since otherwise the complement to its closure would contain an element of the base. $\square$

1.1.3. Comparison of topologies. A topology $\mathcal{S}$ is said to be *stronger* (or *finer*) than $\mathcal{T}$ if $\mathcal{T} \subset \mathcal{S}$, and *weaker* (or *coarser*) if $\mathcal{S} \subset \mathcal{T}$.

There are two extreme topologies on any set: the weakest *trivial topology* with only the whole space and the empty set being open, and the strongest or finest *discrete topology* where all sets are open (and hence closed).

EXAMPLE 1.1.13. On the two point set D, the topology obtained by declaring open (besides D and $\varnothing$) the set consisting of one of the points (but not the other) is strictly finer than the trivial topology and strictly weaker than the discrete topology.

PROPOSITION 1.1.14. *For any set X and any collection $\mathcal{C}$ of subsets of X there exists a unique weakest topology for which all sets from $\mathcal{C}$ are open.*

PROOF. Consider the collection $\mathcal{T}$ which consist of unions of finite intersections of sets from $\mathcal{C}$ and also includes the whole space and the empty set. By properties (2) and (3) of Definition 1.1.1 in any topology in which sets from $\mathcal{C}$ are open the sets from $\mathcal{T}$ are also open. Collection $\mathcal{T}$ satisfies property (1) of Definition 1.1.1 by definition, and it follows immediately from the properties of unions and intersections that $\mathcal{T}$ satisfies (2) and (3) of Definition 1.1.1. $\square$

Any topology weaker than a separable topology is also separable, since any dense set in a stronger topology is also dense in a weaker one.

EXERCISE 1.1.2. How many topologies are there on the 2–element set and on the 3–element set?

EXERCISE 1.1.3. On the integers $\mathbb{Z}$, consider the *profinite* topology for which open sets are defined as unions (not necessarily finite) of arithmetic progressions (non-constant and infinite in both directions). Prove that this defines a topology which is neither discrete nor trivial.

EXERCISE 1.1.4. Define *Zariski* topology in the set of real numbers by declaring complements of finite sets to be open. Prove that this defines a topology which is coarser than the standard one. Give an example of a sequence such that all points are its limits.

EXERCISE 1.1.5. On the set $\mathbb{R} \cup \{*\}$, define a topology by declaring open all sets of the form $\{*\} \cup G$, where $G \subset \mathbb{R}$ is open in the standard topology of $\mathbb{R}$.

(a) Show that this is indeed a topology, coarser than the discrete topology on this set.

(b) Give an example of a convergent sequence which has two limits.

1.2. Continuous maps and homeomorphisms

In this section, we study, in the language of topology, the fundamental notion of continuity and define the main equivalence relation between topological spaces – homeomorphism. We can say (in the category theory language) that now, since the objects (topological spaces) have been defined, we are ready to define the corresponding morphisms (continuous maps) and isomorphisms (topological equivalence or homeomorphism). Categorica prefac

1.2.1. Continuous maps. The topological definition of continuity is simpler and more natural than the ε, δ definition familiar from the elementary real analysis course.

DEFINITION 1.2.1. Let $(X, \mathcal{T})$ and $(Y, \mathcal{S})$ be topological spaces. A map $f\colon X \to Y$ is said to be *continuous* if $O \in \mathcal{S}$ implies $f^{-1}(O) \in \mathcal{T}$ (preimages of open sets are open):

f is an *open map* if it is continuous and $O \in \mathcal{T}$ implies $f(O) \in \mathcal{S}$ (images of open sets are open);

f is *continuous at the point* x if for any neigborhood A of $f(x)$ in Y the preimage $f^{-1}(A)$ contains a neighborhood of x.

A function f from a topological space to $\mathbb{R}$ is said to be *upper semicontinuous* if $f^{-1}(-\infty, c) \in \mathcal{T}$ for all $c \in \mathbb{R}$:

lower semicontinuous if $f^{-1}(c, \infty) \in \mathcal{T}$ for $c \in \mathbb{R}$.

EXERCISE 1.2.1. Prove that a map is continuous if and only if it is continuous at every point.

Let Y be a topological space. For any collection $\mathcal{F}$ of maps from a set X (without a topology) to Y there exists a unique weakest topology on

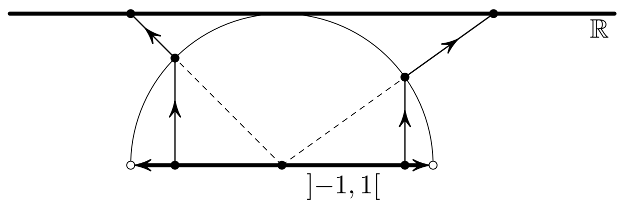

FIGURE 1.2.1. The open interval is homeomorphic to the real line

X which makes all maps from $\mathcal{F}$ continuous; this is exactly the weakest topology with respect to which preimages of all open sets in Y under the maps from $\mathcal{F}$ are open. If $\mathcal{F}$ consists of a single map f, this topology is sometimes called the *pullback topology* on X under the map f.

EXERCISE 1.2.2. Let p be the orthogonal projection of the square K on one of its sides. Describe the pullback topology on K. Will an open (in the usual sense) disk inside K be an open set in this topology?

1.2.2. Topological equivalence. Just as algebraists study groups up to isomorphism or matrices up to a linear conjugacy, topologists study (topological) spaces up to homeomorphism.

DEFINITION 1.2.2. A map $f : X \to Y$ between topological spaces is a *homeomorphism* if it is continuous and bijective with continuous inverse.

If there is a homeomorphism $X \to Y$, then X and Y are said to be *homeomorphic* or sometimes *topologically equivalent*.

A property of a topological space that is the same for any two homeomorphic spaces is said to be a *topological invariant* .

The relation of being homeomorphic is obviously an equivalence relation (in the technical sense: it is reflexive, symmetric, and transitive). Thus topological spaces split into equivalence classes, sometimes called *homeomorphy classes*. In this connection, the topologist is sometimes described as a person who cannot distinguish a coffee cup from a doughnut (since these two objects are homeomorphic). In other words, two homeomorphic topological spaces are identical or indistinguishable from the intrinsic point of view in the same sense as isomorphic groups are indistinguishable from the point of view of abstract group theory or two conjugate $n \times n$ matrices are indistinguishable as linear transformations of an n-dimensional vector space without a fixed basis.

a problem with
his figure in the
page

EXAMPLE 1.2.3. The figure shows how to construct homeomorphisms between the open interval and the open half-circle and between the open half-circle and the real line $\mathbb{R}$, thus establishing that the open interval is homeomorphic to the real line.

EXERCISE 1.2.3. Prove that the sphere $\mathbb{S}^2$ with one point removed is homeomorphic to the plane $\mathbb{R}^2$.

EXERCISE 1.2.4. Prove that any open ball is homeomorphic to $\mathbb{R}^3$.

EXERCISE 1.2.5. Describe a topology on the set $\mathbb{R}^2 \cup \{*\}$ which will make it homeomorphic to the sphere $\mathbb{S}^2$.

To show that certain spaces are homeomorphic one needs to exhibit a homeomorphism; the exercises above give basic but important examples of homeomorphic spaces; we will see many more examples already in the course of this chapter. On the other hand, in order to show that topological spaces are not homeomorphic one need to find an invariant which distinguishes them. Let us consider a very basic example which can be treated with tools from elementary real analysis.

EXAMPLE 1.2.4. In order to show that closed interval is not homeomorphic to an open interval (and hence by Example 1.2.3 to the real line) notice the following. Both closed and open interval as topological spaces have the property that the only sets which are open and closed at the same time are the space itself and the empty set. This follows from characterization of open subsets on the line as finite or countable unions of disjoint open intervals and the corresponding characterization of open subsets of a closed interval as unions of open intervals and semi-open intervals containing endpoints. Now if one takes any point away from an open interval the resulting space with induced topology (see below) will have two proper subsets which are open and closed simultaneously while in the closed (or semi-open) interval removing an endpoint leaves the space which still has no non-trivial subsets which are closed and open.

In Section 1.6 we will develop some of the ideas which appeared in this simple argument systematically.

The same argument can be used to show that the real line $\mathbb{R}$ is not homeomorphic to Euclidean space $\mathbb{R}^n$ for $n \geq 2$ (see Exercise 1.10.7). It is not sufficient however for proving that $\mathbb{R}^2$ is not homeomorphic $\mathbb{R}^3$. Nevertheless, we feel that we intuitively understand the basic structure of the space $\mathbb{R}^n$ and that topological spaces which locally look like $\mathbb{R}^n$ (they are called (n-dimensional) *topological manifolds*) are natural objects of study in topology. Various examples of topological manifolds will appear in the course of this chapter and in Section 1.8 we will introduce precise definitions and deduce some basic properties of topological manifolds.

1.3. Basic constructions

1.3.1. Induced topology. If $Y \subset X$, then Y can be made into a topological space in a natural way by taking the *induced topology*

$$\mathcal{T}_Y := \{O \cap Y \mid O \in \mathcal{T}\}.$$

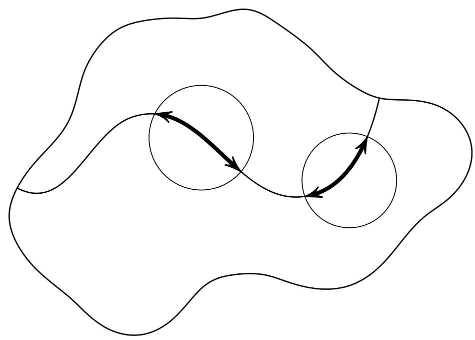

FIGURE 1.3.1. Induced topology

EXAMPLE 1.3.1. The topology induced from $\mathbb{R}^{n+1}$ on the subset

$$\{(x_1, \dots, x_n, x_{n+1}) : \sum_{i=1}^{n+1} x_i^2 = 1\}$$

produces the (standard, or unit) *n–sphere* $\mathbb{S}^n$. For $n = 1$ it is called the (*unit*) *circle* and is sometimes also denoted by $\mathbb{T}$.

EXERCISE 1.3.1. Prove that the boundary of the square is homeomorphic to the circle.

EXERCISE 1.3.2. Prove that the sphere $\mathbb{S}^2$ with any two points removed is homeomorphic to the infinite cylinder $C := \{(x, y, z) \in \mathbb{R}^3 | x^2 + y^2 = 1\}$.

EXERCISE 1.3.3. Let $S := \{(x, y, z) \in \mathbb{R}^3 \mid z = 0,\ x^2 + y^2 = 1\}$. Show that $\mathbb{R}^3 \setminus S$ can be mapped continuously onto the circle.

1.3.2. Product topology. If $(X_\alpha, \mathcal{T}_\alpha)$, $\alpha \in A$ are topological spaces and A is any set, then the *product topology* on $\prod_{\alpha \in A} X$ is the topology determined by the base

$$\Big\{ \prod_\alpha O_\alpha \mid O_\alpha \in \mathcal{T}_\alpha,\ O_\alpha \neq X_\alpha \text{ for only finitely many } \alpha \Big\}.$$

EXAMPLE 1.3.2. The standard topology in $\mathbb{R}^n$ coincides with the product topology on the product of n copies of the real line $\mathbb{R}$.

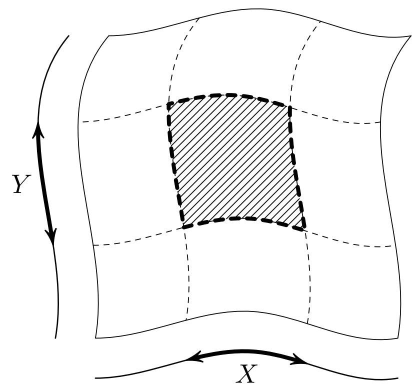

FIGURE 1.3.2. Basis element of the product topology

EXAMPLE 1.3.3. The product of n copies of the circle is called the *n–torus* and is usually denoted by $\mathbb{T}^n$. The n– torus can be naturally identified with the following subset of $\mathbb{R}^{2n}$:

$$\{(x_1, \dots x_{2n}) : x_{2i-1}^2 + x_{2i}^2 = 1,\ i = 1, \dots, n.\}$$

with the induced topology.

EXAMPLE 1.3.4. The product of countably many copies of the two–point space, each with the discrete topology, is one of the representations of the *Cantor set* (see Section 1.7 for a detailed discussion).

EXAMPLE 1.3.5. The product of countably many copies of the closed unit interval is called the *Hilbert cube*. It is the first interesting example of a Hausdorff space (Section 1.4) "too big" to lie inside (that is, to be homeomorphic to a subset of) any Euclidean space $\mathbb{R}^n$. Notice however, that not only we lack means of proving the fact right now but the elementary invariants described later in this chapter are not sufficient for this task either.

1.3.3. Quotient topology. Consider a topological space $(X, \mathcal{T})$ and suppose there is an equivalence relation $\sim$ defined on X. Let π be the natural projection of X on the set $\hat{X}$ of equivalence classes. The *identification space* or *quotient space* $X/\!\sim\, := (\hat{X}, \mathcal{S})$ is the topological space obtained by calling a set $O \subset \hat{X}$ open if $\pi^{-1}(O)$ is open, that is, taking on $\hat{X}$ the finest topology for which π is continuous. For the moment we restrict ourselves to "good" examples, i.e. to the situations where quotient topology is natural in some sense. However the reader should be aware that even very natural equivalence relations often lead to factors with bad properties ranging from the trivial topology to nontrivial ones but lacking basic separation properties (see Section 1.4). We postpone description of such examples till Section 1.9.2.

EXAMPLE 1.3.6. Consider the closed unit interval and the equivalence relation which identifies the endpoints. Other equivalence classes are single points in the interior. The corresponding quotient space is another representation of the circle.

The product of n copies of this quotient space gives another definition of the n–torus.

EXERCISE 1.3.4. Describe the representation of the n–torus from the above example explicitly as the identification space of the unit n–cube I^n:

$$\{(x_1, \ldots, x_n) \in \mathbb{R}^n : 0 \leq x_i \leq 1,\ i = 1, \ldots n.$$

EXAMPLE 1.3.7. Consider the following equivalence relation in punctured Euclidean space $\mathbb{R}^{n+1} \setminus \{0\}$:

$$(x_1, \ldots, x_{n+1}) \sim (y_1, \ldots, y_{n+1}) \text{ iff } y_i = \lambda x_i \text{ for all } i = 1, \ldots, n+1$$

with the same real number λ. The corresponding identification space is called the *real projective n–space* and is denoted by $\mathbb{R}P(n)$.

A similar procedure in which λ has to be positive gives another definition of the n–sphere $\mathbb{S}^n$.

EXAMPLE 1.3.8. Consider the equivalence relation in $\mathbb{C}^{n+1} \setminus \{0\}$:

$$(x_1, \ldots, x_{n+1}) \sim (y_1, \ldots, y_{n+1}) \text{ iff } y_i = \lambda x_i \text{ for all } i = 1, \ldots, n+1$$

with the same complex number λ. The corresponding identification space is called the *complex projective n–space* and is detoted by $\mathbb{C}P(n)$.

EXAMPLE 1.3.9. The map $E : [0,1] \to \mathbb{S}^1, E(x) = \exp 2\pi i x$ establishes a homeomorphism between the interval with identified endpoints (Example 1.3.6) and the unit circle defined in Example 1.3.1.

EXAMPLE 1.3.10. The identification of the equator of the 2-sphere to a point yields two spheres with one common point.

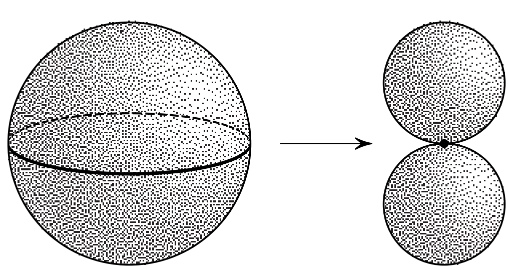

FIGURE 1.3.3. The sphere with equator identified to a point

EXAMPLE 1.3.11. Identifying the short sides of a long rectangle in the natural way yields the lateral surface of the cylinder (which of course is homeomorphic to the annulus), while the identification of the same two sides in the "wrong way" (i.e., after a half twist of the strip) produces the famous Möbius strip. We assume the reader is familiar with the failed experiments of painting the two sides of the Möbius strip in different colors or cutting it into two pieces along its midline. Another less familiar but amusing endeavor is to predict what will happen to the physical object obtained by cutting a paper Möbius strip along its midline if that object is, in its turn, cut along its own midline.

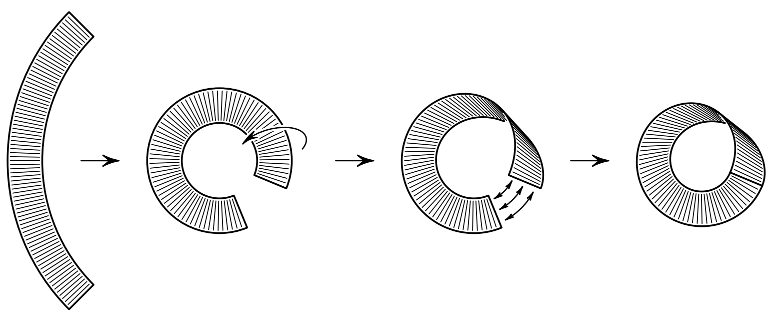

FIGURE 1.3.4. The Möbius strip

EXERCISE 1.3.5. Describe a homeomorphism between the torus $\mathbb{T}^n$ (Example 1.3.3) and the quotient space described in Example 1.3.6 and the subsequent exercise.

EXERCISE 1.3.6. Describe a homeomorphism between the sphere $\mathbb{S}^n$ (Example 1.3.1) and the second quotient space of Example 1.3.7.

EXERCISE 1.3.7. Prove that the real projective space $\mathbb{R}P(n)$ is homeomorphic to the quotient space of the sphere S^n with respect to the equivalence relation which identifies pairs of opposite points: x and $-x$.

EXERCISE 1.3.8. Consider the equivalence relation on the closed unit ball $\mathbb{D}^n$ in $\mathbb{R}^n$:

$$\{(x_1, \dots, x_n) : \sum_{i=1}^{n} x_i^2 \leq 1\}$$

which identifies all points of $\partial\mathbb{D}^n = \mathbb{S}^{n-1}$ and does nothing to interior points. Prove that the quotient space is homeomorphic to $\mathbb{S}^n$.

EXERCISE 1.3.9. Show that $\mathbb{C}P(1)$ is homeomorphic to $\mathbb{S}^2$.

DEFINITION 1.3.12. The *cone* $\operatorname{Cone}(X)$ over a topological space X is the quotient space obtained by identifying all points of the form $(x, 1)$ in the product $(X \times [0, 1]$ (supplied with the product topology).

The *suspension* $\Sigma(X)$ of a topological space X is the quotient space of the product $X \times [-1, 1]$ obtained by identifying all points of the form $x \times 1$ and identifying all points of the form $x \times -1$. By convention, the suspension of the empty set will be the two-point set $\mathbb{S}^0$.

The *join* $X * Y$ of two topological spaces X and Y , roughly speaking, is obtained by joining all pairs of points (x, y), $x \in X$, $y \in Y$, by line segments and supplying the result with the natural topology; more precisele, $X * Y$ is the quotient space of the product $X \times [-1, 1] \times Y$ under the following identifications:

$(x, -1, y) \sim (x, -1, y')$ for any $x \in X$ and all $y, y' \in Y$,
$(x, 1, y) \sim (x', 1, y)$ for any $y \in Y$ and all $x, x' \in X$.

EXAMPLE 1.3.13. (a) $\operatorname{Cone}(*) = \mathbb{D}^1$ and $\operatorname{Cone}(\mathbb{D}^{n-1}) = \mathbb{D}^n$ for $n > 1$.
(b) The suspension $\Sigma(\mathbb{S}^n)$ of the n-sphere is the $(n + 1)$-sphere $\mathbb{S}^{n+1}$.
(c) The join of two closed intervals is the 3-simplex (see the figure).

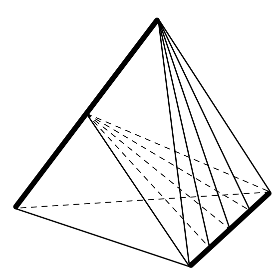

FIGURE 1.3.5. The 3-simplex as the join of two segments

EXERCISE 1.3.10. Show that the cone over the sphere $\mathbb{S}^n$ is (homeomorphic to) the disk $\mathbb{D}^{n+1}$.

EXERCISE 1.3.11. Show that the join of two spheres $\mathbb{S}^k$ and $\mathbb{S}^l$ is (homeomorphic to) the sphere $\mathbb{S}^{k+l+1}$.

EXERCISE 1.3.12. Is the join operation on topological spaces associative?

1.4. Separation properties

Separation properties provide one of the approaches to measuring how fine is a given topology.

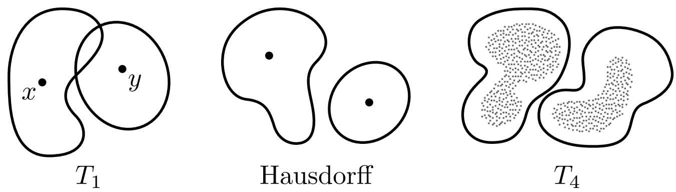

FIGURE 1.4.1. Separation properties

1.4.1. T1, Hausdorff, and normal spaces. Here we list, in decreasing order of generality, the most common separation axioms of topological spaces.

DEFINITION 1.4.1. A topological space $(X, \mathcal{T})$ is said to be a

(T1) *space* if any point is a closed set. Equivalently, for any pair of points $x_1, x_2 \in X$ there exists a neighborhood of x_1 not containing x_2;

(T2) or *Hausdorff space* if any two distinct points possess nonintersecting neighborhoods;

(T4) or *normal space* if it is Hausdorff and any two closed disjoint subsets possess nonintersecting neighborhoods. [1]

It follows immediately from the definition of induced topology that any of the above separation properties is inherited by the induced topology on any subset.

On the picture
does not looks
the e

EXERCISE 1.4.1. Prove that in a (T2) space any sequence has no more than one limit. Show that without the (T2) condition this is no longer true.

EXERCISE 1.4.2. Prove that the product of two (T1) (respectively Hausdorff) spaces is a (T1) (resp. Hausdorff) space.

REMARK 1.4.2. We will see later (Section 1.9) that even very naturally defined equivalence relations in nice spaces may produce quotient spaces with widely varying separation properties.

The word "normal" may be understood in its everyday sense like "commonplace" as in "a normal person". Indeed, normal topological possess many properties which one would expect form commonplaces notions of continuity. Here is an examples of such property dealing with extension of maps:

THEOREM 1.4.3. [Tietze] *If X is a normal topological space, $Y \subset X$ is closed, and $f\colon Y \to [-1, 1]$ is continuous, then there is a continuous*

[1] Hausdorff (or (T1)) assumption is needed to ensure that there are enough closed sets; specifically that points are closed sets. Otherwise trivial topology would satisfy this property.

extension of f to X, i.e., a continuous map $F : X \to [-1, 1]$ such that $F|_Y = f$.

The proof is based on the following fundamental result, traditionally called Urysohn Lemma, which asserts existence of many continuous maps from a normal space to the real line and thus provided a basis for introducing measurements in normal topological spaces (see Theorem 3.5.1) and hence by Corollary 3.5.3 also in compact Hausdorff spaces.

THEOREM 1.4.4. [Urysohn Lemma] *If X is a normal topological space and A, B are closed subsets of X, then there exists a continuous map $u : X \to [0, 1]$ such that $u(A) = \{0\}$ and $u(B) = \{1\}$.*

PROOF. Let V be en open subset of X and U any subset of X such that $\overline{U} \subset V$. Then there exists an open set W for which $\overline{U} \subset W \subset \overline{W} \subset V$. Indeed, for W we can take any open set containing $\overline{U}$ and not intersecting an open neighborhood of $X \setminus V$ (such a W exists because X is normal).

Applying this to the sets $U := A$ and $V := X \setminus B$, we obtain an "intermediate" open set A_1 such that

$$A \subset A_1 \subset X \setminus B, \tag{1.4.1}$$

where $\overline{A_1} \subset X \setminus B$. Then we can introduce the next intermediate open sets A_1' and A_2 so as to have

$$A \subset A_1' \subset A_1 \subset A_2 \subset X \setminus B, \tag{1.4.2}$$

where each set is contained, together with its closure, in the next one.

For the sequence (1.4.1), we define a function $u_1 : X \to [0, 1]$ by setting

$$u_1(x) = \begin{cases} 0 & \text{for } x \in A\ , \\ 1/2 & \text{for } x \in A_1 \setminus A, \\ 1 & \text{for } X \setminus A_1. \end{cases}$$

For the sequence (1.4.2), we define a function $u_2 : X \to [0, 1]$ by setting

$$u_2(x) = \begin{cases} 0 & \text{for } x \in A\ , \\ 1/4' & \text{for } x \in A_1' \setminus A\ , \\ 1/2 & \text{for } x \in A_1 \setminus A_1', \\ 3/4 & \text{for } x \in A_2 \setminus A_1, \\ 1 & \text{for } x \in X \setminus A_2. \end{cases}$$

Then we construct a third sequence by inserting intermediate open sets in the sequence (1.4.2) and define a similar function u_3 for this sequence, and so on. maybe inse

Obviously, $u_2(x) \geq u_1(x)$ for all $x \in X$. Similarly, for any $n > 1$ we have $u_{n+1}(x) \geq u_n(x)$ for all $x \in X$, and therefore the limit function $u(x) := \lim_{n \to infty} u_n(x)$ exists. It only remains to prove that u is continuous.

Suppose that at the nth step we have constructed the nested sequence of sets corresponding to the function u_n

$$A \subset A_1 \subset \dots A_r \subset X \setminus B,$$

where $\overline{A_i} \subset A_{i+1}$. Let $A_0 := \operatorname{int} A$ be the interior of A, let $A_{-1} := \varnothing$, and $A_{r+1} := X$. Consider the open sets $A_{i+1} \setminus \overline{A_{i-1}}$, $i = 0, 1, \dots, r$. Clearly,

$$X = \bigcup_{i=0}^{r} (\bar{A}_i \setminus \overline{A_{i-1}}) \subset \bigcup_{i=0}^{r} (A_{i+1} \setminus \overline{A_{i-1}}),$$

so that the open sets $A_{i+1} \setminus \overline{A_{i-1}}$ cover the entire space X.

On each set $A_{i+1} \setminus \overline{A_{i-1}}$ the function takes two values that differ by $1/2^n$. Obviously,

$$|u(x) - u_n(x)| \leq \sum_{k=n+1}^{\infty} 1/2^k \;=\; 1/2^n.$$

For each point $x \in X$ let us choose an open neighborhood of the form $A_{i+1} \setminus \overline{A_{i-1}}$. The image of the open set $A_{i+1} \setminus \overline{A_{i-1}}$ is contained in the interval $(u(x) - \varepsilon, u(x) + \varepsilon)$, where $\varepsilon < 1/2^n$. Taking $\varepsilon \to \infty$, we see that u is continuous. □

Now let us deduce Theorem 1.4.3 from the Urysohn lemma.

To this end, we put

$$r_k \;:=\; \frac{1}{2}\Big(\frac{2}{3}\Big)^k, \quad k = 1, 2, \dots.$$

Let us construct a sequence of functions $f_1, f_2, \dots$ on X and a sequence of functions $g_1, g_2, \dots$ on Y by induction. First, we put $f_1 := f$. Suppose that the functions $f_1, \dots, f_k$ have been constructed. Consider the two closed disjoint sets

$$A_k := \{x \in X \mid f_k(x) \leq -r_k\} \quad \text{and} \quad B_k := \{x \in X \mid f_k(x) \geq r_k\}.$$

Applying the Urysohn lemma to these sets, we obtain a continuous map $g_k : Y \to [-r_k, r_k]$ for which $g_k(A_k) = \{-r_k\}$ and $g_k(B_k) = \{r_k\}$. On the set A_k, the functions f_k and g_k take values in the interval $]-3r_k, -r_k[$; on the set A_k, they take values in the interval $]r_k, 3r_k[$; at all other points of the set X, these functions take values in the interval $]-r_k, r_k[$.

Now let us put $f_{k+1} := f_k - g_k|_X$. The function f_{k+1} is obviously continuous on X and $|f_{k+1}(x)| \le 2r_k = 3r_{k+1}$ for all $x \in X$.

Consider the sequence of functions $g_1, g_2, \dots$ on Y. By construction, $|g_k(y)| \le r_k$ for all $y \in Y$. The series

$$\sum_{k=1}^{\infty} r_k = \frac{1}{2} \sum_{k=1}^{\infty} \left(\frac{2}{3}\right)^k$$

converges, and so the series $\Sigma_{k=1}^{\infty} g_k(x)$ converges uniformly on Y to some continuous function

$$F(x) := \sum_{k=1}^{\infty} g_k(x).$$

Further, we have

$$(g_1 + \cdots + g_k) = (f_1 - f_2) + (f_2 - f_3) + \cdots + (f_k - f_{k+1}) = f_1 - f_{k+1} = f - f_{k+1}.$$

But $\lim_{k\to\infty} f_{k+1}(y) = 0$ for any $y \in Y$, hence $F(x) = f(x)$ for any $x \in X$, so that F is a continuous extension of f.

It remains to show that $|F(x)| \le 1$. We have

$$\begin{aligned} |F(x)| &\le \sum_{k=1}^{\infty} |g_k(x)| \le \sum_{k=1}^{\infty} r_k = \sum_{k=1}^{\infty} \left(\frac{2}{3}\right)^k \\ &= \sum_{k=1}^{\infty} \left(\frac{2}{3}\right)^k = \frac{1}{3}\left(1 - \frac{2}{3}\right)^{-1} = 1. \qquad \square \end{aligned}$$

COROLLARY 1.4.5. *Let $X \subset Y$ be a closed subset of a normal space Y and let $f : X \to \mathbb{R}$ be continuous. Then f has a continuous extension $F : Y \to \mathbb{R}$.*

PROOF. The statement follows from the Tietze theorem and the Urysohn lemma by appropriately using the rescaling homeomorphism

$$g : \mathbb{R} \to (-\pi/2, \pi/2) \quad \text{given by} \quad g(x) := \arctan(x). \quad \square$$

Most natural topological spaces which appear in analysis and geometry (but not in some branches of algebra) are normal. The most important instance of non-normal topology is discussed in the next subsection.

1.4.2. Zariski topology. The topology that we will now introduce and seems pathological in several aspects (it is non-Hausdorff and does not possess a countable base), but very useful in applications, in particular in algebraic geometry. We begin with the simplest case which was already mentioned in Example 1.1.4

DEFINITION 1.4.6. The *Zariski topology* on the real line $\mathbb{R}$ is defined as the family $\mathcal{Z}$ of all complements to finite sets.

PROPOSITION 1.4.7. *The Zariski topology given above endows* $\mathbb{R}$ *with the structure of a topological space* $(\mathbb{R}, \mathcal{Z})$*, which possesses the following properties:*

(1) *it is a* (T1) *space;*
(2) *it is separable;*
(3) *it is not a Hausdorff space;*
(4) *it does not have a countable base.*

PROOF. All four assertions are fairly straightforward:

(1) the Zariski topology on the real line is (T1), because the complement to any point is open;

(2) it is separable, since it is weaker than the standard topology in $\mathbb{R}$;

(3) it is not Hausdorff, because any two nonempty open sets have nonempty intersection;

(4) it does not have a countable base, because the intersection of all the sets in any countable collection of open sets is nonemply and thus the complement to any point in that intersection does not contain any element from that collection. □

The definition of Zariski topology on $\mathbb{R}$ (Definition 1.4.6) can be straightforwardly generalized to $\mathbb{R}^n$ for any $n \geq 2$, and the assertions of the proposition above remain true. However, this definition is not the natural one, because it generalizes the "wrong form" of the notion of Zariski topology. The "correct form" of that notion originally appeared in algebraic geometry (which studies zero sets of polynomials) and simply says that closed sets in the Zariski topology on $\mathbb{R}$ are sets of zeros of polynomials $p(x) \in \mathbb{R}[x]$. This motivates the following definitions.

DEFINITION 1.4.8. The *Zariski topology* is defined

- in Euclidean space $\mathbb{R}^n$ by stipulating that the sets of zeros of all polynomials are closed;
- on the unit sphere $S^n \subset \mathbb{R}^{n+1}$ by taking for closed sets the sets of zeros of homogeneous polynomials in $n + 1$ variables;
- on the real and complex projective spaces $\mathbb{R}P(n)$ and $\mathbb{C}P(n)$ (Example 1.3.7, Example 1.3.8) via zero sets of homogeneous polynomials in $n + 1$ real and complex variables respectively.

EXERCISE 1.4.3. Verify that the above definitions supply each of the sets $\mathbb{R}^n$, S^n, $\mathbb{R}P(n)$, and $\mathbb{C}P(n)$ with the structure of a topological space satisfying the assertions of Proposition 1.4.7.

1.5. Compactness

The fundamental notion of compactness, familiar from the elementary real analysis course for subsets of the real line $\mathbb{R}$ or of Euclidean space $\mathbb{R}^n$, is defined below in the most general topological situation.

1.5.1. Types of compactness. A family of open sets $\{O_\alpha\} \subset \mathcal{T}, \alpha \in A$ is called an *open cover* of a topological space X if $X = \bigcup_{\alpha \in A} O_\alpha$, and is a finite open cover if A is finite.

DEFINITION 1.5.1. The space $(X, \mathcal{T})$ is called
• *compact* if every open cover of X has a finite subcover;
• *sequentially compact* if every sequence has a convergent subsequence;
• *σ–compact* if it is the union of a countable family of compact sets.
• *locally compact* if every point has an open neighborhood whose closure is compact in the induced topology.

It is known from elementary real analysis that for subsets of a $\mathbb{R}^n$ compactness and sequential compactness are equivalent. This fact naturally generalizes to metric spaces (see Proposition 3.6.4).

PROPOSITION 1.5.2. *Any closed subset of a compact set is compact.*

PROOF. If K is compact, $C \subset K$ is closed, and Γ is an open cover for C, then $\Gamma_0 := \Gamma \cup \{K \setminus C\}$ is an open cover for K, hence Γ_0 contains a finite subcover $\Gamma' \cup \{K \setminus C\}$ for K; therefore Γ' is a finite subcover (of Γ) for C. □

PROPOSITION 1.5.3. *Any compact subset of a Hausdorff space is closed.*

PROOF. Let X be Hausdorff and let $C \subset X$ be compact. Fix a point $x \in X \setminus C$ and for each $y \in C$ take neighborhoods U_y of y and V_y of x such that $U_y \cap V_y = \varnothing$. Then $\bigcup_{y \in C} U_y \supset C$ is a cover of C and has a finite subcover $\{U_{x_i} \mid 0 \le i \le n\}$. Hence $N_x := \bigcap_{i=0}^{n} V_{y_i}$ is a neighborhood of x disjoint from C. Thus

$$X \setminus C = \bigcup_{x \in X \setminus C} N_x$$

is open and therefore C is closed. □

PROPOSITION 1.5.4. *Any compact Hausdorff space is normal.*

PROOF. First we show that a closed set K and a point $p \notin K$ can be separated by open sets. For $x \in K$ there are open sets O_x, U_x such that $x \in O_x$, $p \in U_x$ and $O_x \cap U_x = \varnothing$. Since K is compact, there is a finite subcover $O := \bigcup_{i=1}^{n} O_{x_i} \supset K$, and $U := \bigcap_{i=1}^{n} U_{x_i}$ is an open set containing p disjoint from O.

Now suppose K, L are closed sets. For $p \in L$, consider open disjoint sets $O_p \supset K$, $U_p \ni p$. By the compactness of L, there is a finite subcover $U := \bigcup_{j=1}^{m} U_{p_j} \supset L$, and so $O := \bigcap_{j=1}^{m} O_{p_j} \supset K$ is an open set disjoint from $U \supset L$. $\square$

DEFINITION 1.5.5. A collection of sets is said to have the *finite intersection property* if every finite subcollection has nonempty intersection.

PROPOSITION 1.5.6. *Any collection of compact sets with the finite intersection property has a nonempty intersection.*

PROOF. It suffices to show that in a compact space every collection of closed sets with the finite intersection property has nonempty intersection. Arguing by contradiction, suppose there is a collection of closed subsets in a compact space K with empty intersection. Then their complements form an open cover of K. Since it has a finite subcover, the finite intersection property does not hold. $\square$

EXERCISE 1.5.1. Show that if the compactness assumption in the previous proposition is omitted, then its assertion is no longer true.

EXERCISE 1.5.2. Prove that a subset of $\mathbb{R}$ or of $\mathbb{R}^n$ is compact iff it is closed and bounded.

1.5.2. Compactifications of non-compact spaces.

DEFINITION 1.5.7. A compact topological space K is called a *compactification* of a Hausdorff space $(X, \mathcal{T})$ if K contains a dense subset homeomorphic to X.

The simplest example of compactification is the following.

DEFINITION 1.5.8. The *one-point compactification* of a noncompact Hausdorff space $(X, \mathcal{T})$ is $\hat{X} := (X \cup \{\infty\}, \mathcal{S})$, where

$$\mathcal{S} := \mathcal{T} \cup \{(X \cup \{\infty\}) \smallsetminus K \mid K \subset X \text{ compact}\}.$$

EXERCISE 1.5.3. Show that the one-point compactification of a Hausdorff space X is a compact (T1) space with X as a dense subset. Find a necessary and sufficient condition on X which makes the one-point compactification Hausdorff.

EXERCISE 1.5.4. Describe the one-point compactification of $\mathbb{R}^n$.

Other compactifications are even more important.

EXAMPLE 1.5.9. Real projective space $\mathbb{R}P(n)$ is a compactification of the Euclidean space $\mathbb{R}^n$. This follows easily form the description of $\mathbb{R}P(n)$ as the identification space of a (say, northern) hemisphere with pairs of opposite equatorial points identified. The open hemisphere is homeomorphic to $\mathbb{R}^n$ and the attached "set at infinity" is homeomorphic to the projective space $\mathbb{R}P(n-1)$.

EXERCISE 1.5.5. Describe the complex projective space $\mathbb{C}P(n)$ (see Example 1.3.8) as a compactification of the space $\mathbb{C}^n$ (which is of course homeomorphic to $\mathbb{R}^{2n}$). Specifically, identify the set of added "points at infinity" as a topological space. and desribe open sets which contain points at infinity.

1.5.3. Compactness under products, maps, and bijections. The following result has numerous applications in analysis, PDE, and other mathematical disciplines.

THEOREM 1.5.10. *The product of any family of compact spaces is compact.*

PROOF. Consider an open cover $\mathcal{C}$ of the product of two compact topological spaces X and Y. Since any open neighborhood of any point contains the product of opens subsets in x and Y we can assume that every element of $\mathcal{C}$ is the product of open subsets in X and Y. Since for each $x \in X$ the subset $\{x\} \times Y$ in the induced topology is homeomorphic to Y and hence compact, one can find a finite subcollection $\mathcal{C}_x \subset \mathcal{C}$ which covers $\{x\} \times Y$.

For $(x, y) \in X \times Y$, denote by p_1 the projection on the first factor: $p_1(x, y) = x$. Let $U_x = \bigcap_{C \in \mathcal{O}_x} p_1(C)$; this is an open neighborhood of x and since the elements of $\mathcal{O}_x$ are products, $\mathcal{O}_x$ covers $U_x \times Y$. The sets $\mathcal{U}_x$, $x \in X$ form an open cover of X. By the compactness of X, there is a finite subcover, say $\{U_{x_1}, \dots, U_{x_k}\}$. Then the union of collections $\mathcal{O}_{x_1}, \dots, \mathcal{O}_{x_k}$ form a finite open cover of $X \times Y$.

For a finite number of factors, the theorem follows by induction from the associativity of the product operation and the case of two factors. The proof for an arbitrary number of factors uses some general set theory tools based on axiom of choice. □

PROPOSITION 1.5.11. *The image of a compact set under a continuous map is compact.*

PROOF. If C is compact and $f\colon C \to Y$ continuous and surjective, then any open cover Γ of Y induces an open cover $f_*\Gamma := \{f^{-1}(O) \mid O \in \Gamma\}$ of C which by compactness has a finite subcover $\{f^{-1}(O_i) \mid i = 1, \dots, n\}$. By surjectivity, $\{O_i\}_{i=1}^n$ is a cover for Y. □

A useful application of the notions of continuity, compactness, and separation is the following simple but fundamental result, sometimes referred to as *invariance of domain*:

PROPOSITION 1.5.12. *A continuous bijection from a compact space to a Hausdorff space is a homeomorphism.*

PROOF. Suppose X is compact, Y Hausdorff, $f\colon X \to Y$ bijective and continuous, and $O \subset X$ open. Then $C := X \setminus O$ is closed, hence compact, and $f(C)$ is compact, hence closed, so $f(O) = Y \setminus f(C)$ (by bijectivity) is open. $\square$

Using Proposition 1.5.4 we obtain

COROLLARY 1.5.13. *Under the assumption of Proposition 1.5.12 spaces X and Y are normal.*

EXERCISE 1.5.6. Show that for noncompact X the assertion of Proposition 1.5.12 no longer holds.

1.6. Connectedness and path connectedness

There are two rival formal definitions of the intuitive notion of connectedness of a topological space. The first is based on the idea that such a space "consists of one piece" (i.e., does not "fall apart into two pieces"), the second interprets connectedness as the possibility of "moving continuously from any point to any other point".

1.6.1. Definition and invariance under continuous maps.

DEFINITION 1.6.1. A topological space $(X, \mathcal{T})$ is said to be

• *connected* if X cannot be represented as the union of two nonempty disjoint open sets (or, equivalently, two nonempty disjoint closed sets);

• *path connected* if for any two points $x_0, x_1 \in X$ there exists a path joining x_0 to x_1, i.e., a continuous map $c\colon [0,1] \to X$ such that $c(i) = x_i,\ i = \{0,1\}$.

PROPOSITION 1.6.2. *The continuous image of a connected space X is connected.*

PROOF. If the image is decomposed into the union of two disjoint open sets, the preimages of theses sets which are open by continuity would give a similar decomposition for X. $\square$

PROPOSITION 1.6.3. (1) *Interval is connected*

(2) *Any path-connected space is connected.*

PROOF. (1) Any open subset X of an interval is the union of disjoint open subintervals. The complement of X contains the endpoints of those intervals and hence cannot be open.

(2) Suppose X is path-connected and let $x = X_0 \cup X_1$, where X_0 and X_1 are open and nonempty. Let $x_0 \in X_0,\ x_1 \in X_1$ and $c\colon [0,1] \to X$ is a continuous map such that $c(i) = x_i,\ i \in \{0,1\}$. By Proposition 1.6.2 the image $c([0,1])$ is a connected subset of X in induced topology which is decomposed into the union of two nonempty open subsets $c([0,1]) \cap X_0$ and $c([0,1]) \cap X_1$, a contradiction. □

REMARK 1.6.4. Connected space may not be path-connected as is shown by the union of the graph of $\sin 1/x$ and $\{0\} \times [-1,1]$ in $\mathbb{R}^2$ (see the figure).

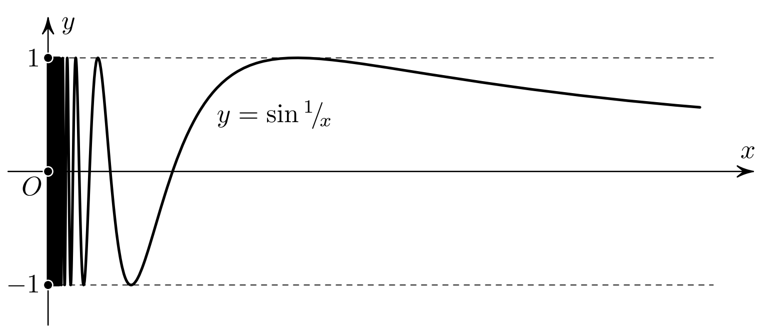

FIGURE 1.6.1. Connected but not path connected space

PROPOSITION 1.6.5. *The continuous image of a path connected space X is path connected.*

PROOF. Let $f : X \to Y$ be continuous and surjective; take any two points $y_1, y_2 \in Y$. Then by surjectivity the sets $f^{-1}(y_i)$, $i = 1,2$ are nonempty and we can choose points $x_i \in f^{-1}(y_1)$, $i = 1,2$. Since X is path connected, there is a path $\alpha : [0,1] \to X$ joining x_1 to x_2. But then the path $f \circ \alpha$ joins y_1 to y_2. □

1.6.2. Products and quotients.

PROPOSITION 1.6.6. *The product of two connected topological spaces is connected.*

PROOF. Suppose $X,\ Y$ are connected and assume that $X \times Y = A \cup B$, where A and B are open, and $A \cap B = \varnothing$. Then either $A = X_1 \times Y$ for some open $X_1 \subset X$ or there exists an $x \in X$ such that $\{x\} \times Y \cap A \neq \varnothing$ and $\{x\} \times Y \cap B \neq \varnothing$.

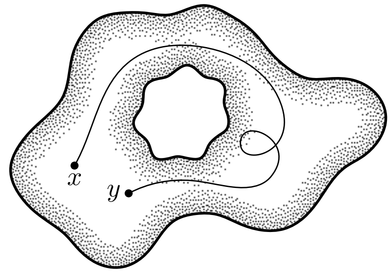

FIGURE 1.6.2. Path connectedness

The former case is impossible, else we would have $B = (X \setminus X_1) \times Y$ and so $X = X_1 \cup (X \setminus X_1)$ would not be connected.

In the latter case, $Y = p_2(\{x\} \times Y \cap A) \cup p_2(\{x\} \times Y \cap B)$ (where $p_2(x, y) = y$ is the projection on the second factor) that is, $\{x\} \times Y$ is the union of two disjoint open sets, hence not connected. Obviously p_2 restricted to $\{x\}\times Y$ is a homeomorphism onto Y, and so Y is not connected either, a contradiction. □

PROPOSITION 1.6.7. *The product of two path-connected topological spaces is connected.*

PROOF. Let $(x_0, y_0),\ (x_1, y_1) \in X \times Y$ and $c_X,\ c_Y$ are paths connecting x_0 with x_1 and y_0 with y_1 correspondingly. Then the path $c\colon [0, 1] \to X \times Y$ defined by

$$c(t) = (c_X(t), c_Y(t))$$

connects (x_0, y_0) with (x_1, y_1). □

The following property follows immediately from the definition of the quotient topology

PROPOSITION 1.6.8. *Any quotient space of a connected topological space is connected.*

1.6.3. Connected subsets and connected components. A subset of a topological space is *connected* (*path connected*) if it is a connected (path connected) space in the induced topology.

A *connected component* of a topological space X is a maximal connected subset of X.

A *path connected component* of X is a maximal path connected subset of X.

PROPOSITION 1.6.9. *The closure of a connected subset $Y \subset X$ is connected.*

PROOF. If $\bar{Y} = Y_1 \cup Y_2$, where $Y_1,\ Y_2$ are open and $Y_1 \cap Y_2 = \varnothing$, then since the set Y is dense in its closure $Y = (Y \cap Y_1) \cup (Y \cap Y_2)$ with both $Y \cap Y_1$ and $Y \cap Y_1$ open in the induced topology and nonempty. □

COROLLARY 1.6.10. *Connected components are closed.*

PROPOSITION 1.6.11. *The union of two connected subsets $Y_1, Y_2 \subset X$ such that $Y_1 \cap Y_2 \neq \varnothing$, is connected.*

PROOF. We will argue by contradiction. Assume that $Y_1 \cap Y_2$ is the disjoint union of of open sets Z_1 and Z_2. If $Z_1 \supset Y_1$, then $Y_2 = Z_2 \cup (Z_1 \cap Y_2)$ and hence Y_2 is not connected. Similarly, it is impossible that $Z_2 \supset Y_1$. Thus $Y_1 \cap Z_i \neq \varnothing,\ i = 1, 2$ and hence $Y_1 = (Y_1 \cap Z_1) \cup (Y_1 \cap Z_2)$ and hence Y_1 is not connected. □

1.6.4. Decomposition into connected components. For any topological space there is a unique *decomposition into connected components* and a unique *decomposition into path connected components*. The elements of these decompositions are equivalence classes of the following two equivalence relations respectively:

(i) x is equivalent to y if there exists a connected subset $Y \subset X$ which contains x and y.

In order to show that the equivalence classes are indeed connected components, one needs to prove that they are connected. For, if A is an equivalence class, assume that $A = A_1 \cup A_2$, where A_1 and A_2 are disjoint and open. Pick $x_1 \in A_1$ and $x_2 \in A_2$ and find a closed connected set A_3 which contains both points. But then $A \subset (A_1 \cup A_3) \cup A_2$, which is connected by Proposition 1.6.11. Hence $A = (A_1 \cup A_3) \cup A_2)$ and A is connected.

(ii) x is equivalent to y if there exists a continuous curve $c\colon [0, 1] \to X$ with $c(0) = x,\ c(1) = y$

REMARK 1.6.12. The closure of a path connected subset may be fail to be path connected. It is easy to construct such a subset by looking at Remark 1.6.4

1.6.5. Arc connectedness. Arc connectedness is a more restrictive notion than path connectedness: a topological space X is called *arc connected* if, for any two distinct points $x, y \in X$ there exist an arc joining them, i.e., there is an injective continuous map $h : [0, 1] \to X$ such that $h(0) = x$ and $h(1) = y$.

It turns out, however, that arc connectedness is not a much more stronger requirement than path connectedness – in fact the two notions coincide for Hausdorff spaces.

THEOREM 1.6.13. *A Hausdorff space is arc connected if and only if it is path connected.*

PROOF. Let X be a path-connected Hausdorff space, $x_0, x_1 \in X$ and $c\colon [0,1] \to X$ a continuous map such that $c(i) = x_i,\ i = 0, 1$. Notice that the image $c([0,1])$ is a compact subset of X by Proposition 1.5.11 even though we will not use that directly. We will change the map c within this image by successively cutting of superfluous pieces and rescaling what remains.

Consider the point $c(1/2)$. If it coincides with one of the endpoints x_o or x_1 we define $c_1(t)$ as $c(2t-1)$ or $c(2t)$ correspondingly. Otherwise consider pairs $t_0 < 1/2 < t_1$ such that $c(t_0) = c(t_1)$. The set of all such pairs is closed in the product $[0,1] \times [0,1]$ and the function $|t_0 - t_1|$ reaches maximum on that set. If this maximum is equal to zero the map c is already injective. Otherwise the maximum is positive and is reached at a pair (a_1, b_1). we define the map c_1 as follows

$$c_1(t) = \begin{cases} c(t/2a_1), \ if \ 0 \le t \le a_1, \\ c(1/2), \ if \ a_1 \le t \le b_1, \\ c(t/2(1-b_1) + (1-b_1)/2), \ if\, b_1 \le t \le 1. \end{cases}$$

Notice that $c_1([0,1/2))$ and $c_1((1/2,1])$ are disjoint since otherwise there would exist $a' < a_1 < b_1 < b'$ such that $c(a') = c(b')$ contradicting maximality of the pair (a_1, b_1).

Now we proceed by induction. We assume that a continuous map $c_n\colon [0,1] \to c([0,1])$ has been constructed such that the images of intervals $(k/2^n, (k+1)/2^n),\ k = 0, \dots, 2^n - 1$ are disjoint. Furthermore, while we do not exclude that $c_n(k/2^n) = c_n((k+1)/2^n)$ we assume that $c_n(k/2^n) \neq c_n(l/2^n)$ if $|k-l| > 1$.

We find a_n^k, b_n^k maximizing the difference $|t_0 - t_1|$ among all pairs

$$(t_0, t_1): \ k/2^n \le t_0 \le t_1 \le (k+1)/2^n$$

and construct the map c_{n+1} on each interval $[k/2^n, (k+1)/2^n]$ as above with c_n in place of c and a_n^k, b_n^k in place of a_1, b_1 with the proper renormalization. As before special provision are made if c_n is injective on one of the intervals (in this case we set $c_{n+1} = c_n$) of if the image of the midpoint coincides with that of one of the endpoints (one half is cut off that the other renormalized). $\square$

EXERCISE 1.6.1. Give an example of a path connected but not arc connected topological space.

1.7. Totally disconnected spaces and Cantor sets

On the opposite end from connected spaces are those spaces which do not have any connected nontrivial connected subsets at all.

1.7.1. Examples of totally disconnected spaces.

DEFINITION 1.7.1. A topological space $(X, \mathcal{T})$ is said to be *totally disconnected* if every point is a connected component. In other words, the only connected subsets of a totally disconnected space X are single points.

Discrete topologies (all points are open) give trivial examples of totally disconnected topological spaces. Another example is the set

$$\left\{0, 1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots, \right\}$$

with the topology induced from the real line. More complicated examples of compact totally disconnected space in which isolated points are dense can be easily constructed. For instance, one can consider the set of rational numbers $\mathbb{Q} \subset \mathbb{R}$ with the induced topology (which is not locally compact).

The most fundamental (and famous) example of a totally disconnected set is the Cantor set, which we now define.

DEFINITION 1.7.2. The (standard middle-third) *Cantor set* $C(1/3)$ is defined as follows:

$$C(1/3); = \left\{x \in \mathbb{R} : x = \sum_{i=1}^{\infty} \frac{x_i}{3^i},\ x_i \in \{0, 2\},\ i = 1, 2, \dots\right\}.$$

Geometrically, the construction of the set $C(1/3)$ may be described in the following way: we start with the closed interval $[0, 1]$, divide it into three equal subintervals and throw out the (open) middle one, divide each of the two remaining ones into equal subintervals and throw out the open middle ones and continue this process *ad infinitum*. What will be left? Of course the (countable set of) endpoints of the removed intervals will remain, but there will also be a much larger (uncountable) set of remaining "mysterious points", namely those which do not have the ternary digit 1 in their ternary expansion.

1.7.2. Lebesgue measure of Cantor sets. There are many different ways of constructing subsets of $[0, 1]$ which are homeomorphic to the Cantor set $C(1/3)$. For example, instead of throwing out the middle one third intervals at each step, one can do it on the first step and then throw out intervals of length $\frac{1}{18}$ in the middle of two remaining interval and inductively throw out the interval of length $\frac{1}{2^n 3^{n+1}}$ in the middle of each of 2^n intervals which remain after n steps. Let us denote the resulting set $\hat{C}$

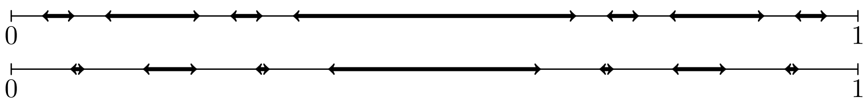

FIGURE 1.7.1. Two Cantor sets

EXERCISE 1.7.1. Prove (by computing the infinite sum of lengths of the deleted intervals) that the Cantor set $C(1/3)$ has Lebesgue measure 0 (which was to be expected), whereas the set $\hat{C}$, although nowhere dense, has *positive* Lebesgue measure.

1.7.3. Some other strange properties of Cantor sets. Cantor sets can be obtained not only as subsets of $[0, 1]$, but in many other ways as well.

PROPOSITION 1.7.3. *The countable product of two point spaces with the discrete topology is homeomorphic to the Cantor set.*

PROOF. To see that, identify each factor in the product with $\{0, 2\}$ and consider the map

$$(x_1, x_2, \dots) \mapsto \sum_{i=1}^{\infty} \frac{x_i}{3^i}, \quad x_i \in \{0, 2\},\ i = 1, 2, \dots .$$

This map is a homeomorphism between the product and the Cantor set. □

PROPOSITION 1.7.4. *The product of two* (*and hence of any finite number*) *of Cantor sets is homeomorphic to the Cantor set.*

PROOF. This follows immediately, since the product of two countable products of two point spaces can be presented as such a product by mixing coordinates. □

EXERCISE 1.7.2. Show that the product of countably many copies of the Cantor set is homeomorphic to the Cantor set.

The Cantor set is a compact Hausdorff with countable base (as a closed subset of $[0, 1]$), and it is *perfect* i.e. has no isolated points. As it turns out, it is a universal model for compact totally disconnected perfect Hausdorff topological spaces with countable base, in the sense that any such space is homeomorphic to the Cantor set $C(1/3)$. This statement will be proved later by using the machinery of metric spaces (see Theorem 3.6.7). For now we restrict ourselves to a certain particular case.

PROPOSITION 1.7.5. *Any compact perfect totally disconnected subset A of the real line $\mathbb{R}$ is homeomorphic to the Cantor set.*

PROOF. The set A is bounded, since it is compact, and nowhere dense (does not contain any interval), since it is totally disconnected. Suppose $m = \inf A$ and $M = \sup A$. We will outline a construction of a strictly monotone function $F : [0,1] \to [m, M]$ such that $F(C) = A$. The set $[m, M] \setminus A$ is the union of countably many disjoint intervals without common ends (since A is perfect). Take one of the intervals whose length is maximal (there are finitely many of them); denote it by I. Define F on the interval I as the increasing linear map whose image is the interval $[1/3, 2/3]$. Consider the longest intervals I_1 and I_2 to the right and to the left to I. Map them linearly onto $[1/9.2/9]$ and $[7/9, 8/9]$, respectively. The complement $[m, M] \setminus (I_1 \cup I \cup I_2)$ consists of four intervals which are mapped linearly onto the middle third intervals of $[0,1] \setminus ([1/9.2/9] \cup [1/3, 2/3] \cup [7/9, 8/9]$ and so on by induction. Eventually one obtains a strictly monotone bijective map $[m, M] \setminus A \to [0,1] \setminus C$ which by continuity is extended to the desired homeomorphism. □

EXERCISE 1.7.3. Prove that the product of countably many finite sets with the discrete topology is homeomorphic to the Cantor set.

1.8. Topological manifolds

At the other end of the scale from totally disconnected spaces are the most important objects of algebraic and differential topology: the spaces which locally look like a Euclidean space. This notion was first mentioned at the end of Section 1.2 and many of the examples which we have seen so far belong to that class. Now we give a rigorous definition and discuss some basic properties of manifolds.

1.8.1. Definition and some properties. The precise definition of a topological manifold is as follows.

DEFINITION 1.8.1. A *topological manifold* is a Hausdorff space X with a countable base for the topology such that every point is contained in an open set homeomorphic to a ball in $\mathbb{R}^n$ for some $n \in \mathbb{N}$. A pair (U, h) consisting of such a neighborhood and a homeomorphism $h \colon U \to B \subset \mathbb{R}^n$ is called a *chart* or a system of *local coordinates*.

e illustrating the definition

REMARK 1.8.2. Hausdorff condition is essential to avoid certain pathologies which we will discuss laler.

Obviously, any open subset of a topological manifold is a topological manifold.

If X is connected, then n is constant. In this case it is called the *dimension* of the topological manifold. Invariance of the dimension (in other words, the fact that $\mathbb{R}^n$ or open sets in those for different n are not homeomorphic) is one of the basic and nontrivial facts of topology.

PROPOSITION 1.8.3. *A connected topological manifold is path connected.*

PROOF. Path connected component of any point in a topological manifold is open since if there is a path from x to y there is also a path from x to any point in a neighborhood of y homeomorphic to $\mathbb{R}^n$. For, one can add to any path the image of an interval connecting y to a point in such a neighborhood. If a path connected component is not the whole space its complement which is the union of path connected components of its points is also open thus contradicting connectedness. $\square$

1.8.2. Examples and constructions.

EXAMPLE 1.8.4. The n–sphere $\mathbb{S}^n$, the n–torus $\mathbb{T}^n$ and the real projective n–space $\mathbb{R}P(n)$ are examples of n dimensional connected topological manifolds; the complex projective n–space $\mathbb{C}P(n)$ is a topological manifold of dimension $2n$.

EXAMPLE 1.8.5. Surfaces in 3-space, i.e., compact connected subsets of $\mathbb{R}^3$ locally defined by smooth functions of two variables x, y in appropriately chosen coordinate systems (x, y, z), are examples of 2-dimensional manifolds.

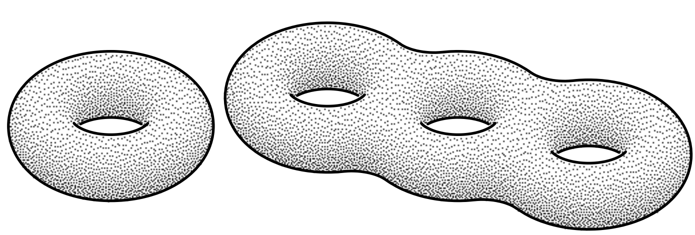

FIGURE 1.8.1. Two 2-dimensional manifolds

EXAMPLE 1.8.6. Let $F : \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function and let c be a noncritical value of F, that is, there are no critical points at which the value of F is equal to c. Then $F^{-1}(c)$ (if nonempty) is a topological manifold of dimension $n - 1$. This can be proven using the Implicit Function theorem from multivariable calculus.

Among the most important examples of manifolds from the point of view of applications, are configuration spaces and phase spaces of mechanical systems (i.e., solid mobile instruments obeying the laws of classical mechanics). One can think of the configuration space of a mechanical system as a topological space whose points are different "positions" of the system, and neighborhoods are "nearby" positions (i.e., positions that can be obtained from the given one by motions of "length" smaller than a fixed number). The phase space of a mechanical system moving in time is obtained from its configuration space by supplying it with all possible velocity vectors. There will be numerous examples of phase and configuration spaces further in the course, here we limit ourselves to some simple illustrations.

EXAMPLE 1.8.7. The configuration space of the mechanical system consisting of a rod rotating in space about a fixed hinge at its extremity is the 2-sphere. If the hinge is fixed at the midpoint of the rod, then the configuration space is $\mathbb{R}P^2$.

EXERCISE 1.8.1. Prove two claims of the previous example.

EXERCISE 1.8.2. The *double pendulum* consists of two rods AB and CD moving in a vertical plane, connected by a hinge joining the extremities B and C, while the extremity A is fixed by a hinge in that plane. Find the configuration space of this mechanical system.

EXERCISE 1.8.3. Show that the configuration space of an asymmetric solid rotating about a fixed hinge in 3-space is $\mathbb{R}P^3$.

EXERCISE 1.8.4. On a round billiard table, a pointlike ball moves with uniform velocity, bouncing off the edge of the table according to the law saying that the angle of incidence is equal to the angle of reflection (see the figure). Find the phase space of this system.

FIGURE ?? Billiards on a circular table

Another source of manifolds with interesting topological properties and usually additional geometric structures is geometry. Spaces of various geometric objects are endowed with a the natural topology which is often generated by a natural metric and also possess natural groups of homeomorphisms.

The simplest non-trival case of this is already familiar.

EXAMPLE 1.8.8. The real projective space $\mathbb{R}P(n)$ has yet another description as the space of all lines in $\mathbb{R}^{n+1}$ passing through the origin. One can define the distance d between two such line as the smallest of four angles between pairs of unit vectors on the line. This distance generates the same topology as the one defined before. Since any invertible linear transformation of $\mathbb{R}^{n+1}$ takes lines into lines and preserves the origin it naturally acts by bijections on $\mathbb{R}P(n)$. Those bijections are homeomorphisms but in general they do not preserve the metric described above or any metric generating the topology.

EXERCISE 1.8.5. Prove claims of the previous example: (i) the distance d defines the same topology on the space $\mathbb{R}^{n+1}$ as the earlier constructions; (ii) the group $GL(n+1, \mathbb{R})$ of invertible linear transformations of $\mathbb{R}^{n+1}$ acts on $\mathbb{R}P(n)$ by homeomorphisms.

There are various modifications and generalizations of this basic example.

EXAMPLE 1.8.9. Consider the space of all lines in the Euclidean plane. Introduce topology into it by declaring that a base of neighborhoods of a given line L consist of the sets $N_L(a, b, \epsilon)$ where $a, b \in L$, $\epsilon > 0$ and $N_L(a, b, \epsilon)$ consist of all lines L' such that the interval of L between a and b lies in the strip of width ϵ around L'

EXERCISE 1.8.6. Prove that this defines a topology which makes the space of lines homeomorphic to the Möbius strip.

EXERCISE 1.8.7. Describe the action of the group $GL(2, \mathbb{R})$ on the Möbius strip coming from the linear action on $\mathbb{R}^2$.

This is the simplest example of the family of *Grassmann manifolds* or *Grassmannians* which play an exceptionally important role in several branches of mathematics including algebraic geometry and theory of group representation. The general Grassmann manifold $G_{k,n}$ (over $\mathbb{R}$) is defined for $i \leq k < n$ as the space of all k-dimensional affine subspaces in $\mathbb{R}^n$. In order to define a topology we again define a base of neighborhoods of a given k-space L. Fix $\epsilon > 0$ and $k+1$ points $x_1, \ldots, x_{k+1} \in L$. A neighborhood of L consists of all k-dimensional spaces L' such that the convex hull of points $x_1, \ldots, x_{k+1}$ lies in the ϵ-neighborhood of L'.

EXERCISE 1.8.8. Prove that the Grassmannian $G_{k,n}$ is a topological manifold. Calculate its dimension. [2]

Another extension deals with replacing $\mathbb{R}$ by $\mathbb{C}$ (and also by quaternions).

EXERCISE 1.8.9. Show that the complex projective space $\mathbb{C}P(n)$ is homeomorphic to the space of all lines on $\mathbb{C}^{n+1}$ with topology defined by a distance similarly to the case of $\mathbb{R}P(n)$

EXERCISE 1.8.10. Define complex Grassmannians, prove that they are manifolds and calculate the dimension.

1.8.3. Additional structures on manifolds. It would seem that the existence of local coordinates should make analysis in $\mathbb{R}^n$ an efficient tool in the study of topological manifolds. This, however, is not the case, because global questions cannot be treated by the differential calculus unless the coordinates in different neighborhoods are connected with each other via *differentiable* coordinate transformations. Notice that continuous functions may be quite pathological form the "normal" commonplace point of view. This requirement leads to the notion of *differentiable manifold*, which will be introduced in Chapter 4 and further studied in Chapter 10. Actually, all the manifolds in the examples above are differentiable, and it has been proved that all manifolds of dimension $n \leq 3$ have a differentiable structure, which is unique in a certain natural sense. For two-dimensional manifolds we will prove this later in Section 5.2.3; the proof for three–dimensional manifolds goes well beyond the scope of this book.

Furthermore, this is no longer true in higher dimensions: there are manifolds that possess no differentiable structure at all, and some that have more than one differentiable structure.

Another way to make topological manifolds more manageable is to endow them with a polyhedral structure, i.e., build them from simple geometric "bricks" which must fit together nicely. The bricks used for this purpose are n-simplices, shown on the figure for $n = 0, 1, 2, 3$ (for the formal definition for any n, see ??).

[2]Remember that we cannot as yet prove that dimension of a connected topological manifold is uniquely defined, i.e. that the same space cannot be a topological manifold of two different dimensions since we do not know that $\mathbb{R}^n$ for different n are not homeomorphic. The question asks to calculate dimension as it appears in the proof that the spaces are manifolds.

FIGURE ?? Simplices of dimension 0,1,2,3.

A PL-structure on an n-manifold M is obtained by representing M as the union of k-simplices, $0 \leq k \leq n$, which intersect pairwise along simplices of smaller dimensions (along "common faces"), and the set of all simplices containing each vertex (0-simplex) has a special "disk structure". This representation is called a *triangulation*. We do not give precise definitions here, because we do not study n-dimensional PL-manifolds in this course, except for $n = 1, 2$, see ?? and ??. In chapter ?? we study a more general class of topological spaces with allow a triangulation, the *simplicial complexes*.

Connections between differentiable and PL structures on manifolds are quite intimate: in dimension 2 existence of a differentiable structure will be derived from simplicial decomposition in **??**. Since each two-dimensional simplex (triangle) possesses the natural smooth structure and in a triangulation these structures in two triangles with a common edge argee along the edge, the only issue here is to "smooth out" the structure around the corners of triangles forming a triangulation.

Conversely, in any dimension any differentiable manifold can be triangulated. The proof while ingenuous uses only fairly basic tools of differential topology.

Again for large values of n not all topological n-manifolds possess a PL-structure, not all PL-manifolds possess a differentiable structure, and when they do, it is not necessarily unique. These are deep and complicated results obtained in the 1970ies, which are way beyond the scope of this book.

1.9. Orbit spaces for group actions

An important class of quotient spaces appears when the equivalence relation is given by the action of a group X by homeomorphisms of a topological space X.

1.9.1. Main definition and nice examples. The notion of a group acting on a space, which formalizes the idea of symmetry, is one of the most important in contemporary mathematics and physics.

DEFINITION 1.9.1. An *action* of a group G on a topological space X is a map $G \times X \to X$, $(g, x) \mapsto xg$ such that

(1) $(xg)h = x(g \cdot h)$ for all $g, h \in G$;

(2) $(x)e = x$ for all $x \in X$, where e is the unit element in G.

The equivalence classes of the corresponding identification are called *orbits* of the action of G on X.

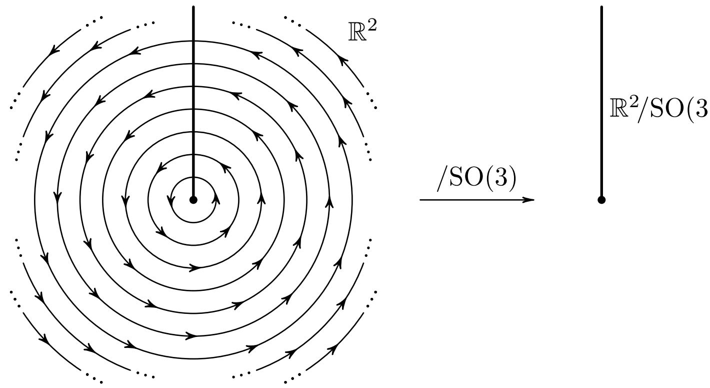

FIGURE 1.9.1. Orbits and identification space of $SO(2)$ action on $\mathbb{R}^2$

The identification space in this case is denoted by X/G and called the *quotient of* X *by* G or the *orbit space* of X under the action of G.

We use the notation xg for the point to which the element g takes the point x, which is more convenient than the notation $g(x)$ (nevertheless, the latter is also often used). To specify the chosen notation, one can say that G acts on X *from the right* (for our notation) or *from the left* (when the notation $g(x)$ or gx is used).

Usually, in the definition of an action of a group G on a space X, the group is supplied with a topological structure and the action itself is assumed continuous. Let us make this more precise.

A *topological group* G is defined as a topological Hausdorff space supplied with a continuous group operation, i.e., an operation such that the maps $(g, h) \mapsto gh$ and $g \mapsto g^{-1}$ are continuous. If G is a finite or countable group, then it is supplied with the discrete topology. When we speak of the action of a topological group G on a space X, we tacitly assume that the map $X \times G \to X$ is a continuous map of topological spaces.

EXAMPLE 1.9.2. Let X be the plane $\mathbb{R}^2$ and G be the rotation group $SO(2)$. Then the orbits are all the circles centered at the origin and the origin itself. The orbit space of $\mathbb{R}^2$ under the action of $SO(2)$ is in a natural bijective correspondence with the half-line $\mathbb{R}_+$.

The main issue in the present section is that in general the quotient space even for a nice looking group acting on a good (for example, locally compact normal with countable base) topological space may not have good separation properties. The (T1) property for the identification space is easy to ascertain: every orbit of the action must be closed. On the other hand, there does not seem to be a natural necessary and sufficient condition for the quotient space to be Hausdorff. Some useful sufficient conditions will appear in the context of metric spaces.

Still, lots of important spaces appear naturally as such identification spaces.

EXAMPLE 1.9.3. Consider the natural action of the integer lattice $\mathbb{Z}^n$ by translations in $\mathbb{R}^n$. The orbit of a point $p \in \mathbb{R}^n$ is the copy of the integer lattice $\mathbb{Z}^n$ translated by the vector p. The quotient space is homeomorphic to the torus $\mathbb{T}^n$.

An even simpler situation produces a very interesting example.

EXAMPLE 1.9.4. Consider the action of the cyclic group of two elements on the sphere S^n generated by the central symmetry: $Ix = -x$. The corresponding quotient space is naturally identified with the real projective space $\mathbb{R}P(n)$.

EXERCISE 1.9.1. Consider the cyclic group of order q generated by the rotation of the circle by the angle $2\pi/q$. Prove that the identification space is homeomorphic to the circle.

EXERCISE 1.9.2. Consider the cyclic group of order q generated by the rotation of the plane $\mathbb{R}^2$ around the origin by the angle $2\pi/q$. Prove that the identification space is homeomorphic to $\mathbb{R}^2$.

1.9.2. Not so nice examples. Here we will see that even simple actions on familiar spaces can produce unpleasant quotients.

EXAMPLE 1.9.5. Consider the following action A of $\mathbb{R}$ on $\mathbb{R}^2$: for $t \in \mathbb{R}$ let $A_t(x, y) = (x+ty, y)$. The orbit space can be identified with the union of two coordinate axis: every point on the x-axis is fixed and every orbit away from it intersects the y-axis at a single point. However the quotient topology is weaker than the topology induced from $\mathbb{R}^2$ would be. Neighborhoods of the points on the y-axis are ordinary but any neighborhood of a point on the x-axis includes a small open interval of the y-axis around the origin. Thus points on the x-axis cannot be separated by open neighborhoods and the space is (T1) (since orbits are closed) but not Hausdorff.

An even weaker but still nontrivial separation property appears in the following example.

EXAMPLE 1.9.6. Consider the action of $\mathbb{Z}$ on $\mathbb{R}$ generated by the map $x \to 2x$. The quotient space can be identified with the union of the circle and an extra point p. Induced topology on the circle is standard. However, the only open set which contains p is the whole space! See Exercise 1.10.21.

Finally let us point out that if all orbits of an action are dense, then the quotient topology is obviously trivial: there are no invariant open sets other than $\varnothing$ and the whole space. Here is a concrete example.

EXAMPLE 1.9.7. Consider the action T of $\mathbb{Q}$, the additive group of rational number on $\mathbb{R}$ by translations: put $T_r(x) = x + r$ for $r \in \mathbb{Q}$ and $x \in \mathbb{R}$. The orbits are translations of $\mathbb{Q}$, hence dense. Thus the quotient topology is trivial.

1.10. Problems

EXERCISE 1.10.1. How many non-homeomorphic topologies are there on the 2–element set and on the 3–element set?

EXERCISE 1.10.2. Let $S := \{(x, y, z) \in \mathbb{R}^3 \,|\, z = 0,\ x^2 + y^2 = 1\}$. Show that $\mathbb{R}^3 \setminus S$ can be mapped continuously onto the circle.

EXERCISE 1.10.3. Consider the product topology on the product of countably many copies of the real line. (this product space is sometimes denoted $\mathbb{R}^\infty$).

a) Does it have a countable base?

b) Is it separable?

EXERCISE 1.10.4. Consider the space $\mathcal{L}$ of all bounded maps $\mathbb{Z} \to \mathbb{Z}$ with the topology of pointwise convergece.

a) Describe the open sets for this topology.

b) Prove that $\mathcal{L}$ is the countable union of disjoint closed subsets each homeomorphic to a Cantor set.

Hint: Use the fact that the countable product of two–point spaces with the product topology is homeomorphic to a Cantor set.

EXERCISE 1.10.5. Consider the *profinite* topology on $\mathbb{Z}$ in which open sets are defined as unions (not necessarily finite) of (non-constant and infinite in both directions) arithmetic progressions. Show that it is Hausdorff but not discrete.

EXERCISE 1.10.6. Let $\mathbb{T}^\infty$ be the product of countably many copies of the circle with the product topology. Define the map $\varphi : \mathbb{Z} \to \mathbb{T}^\infty$ by

$$\varphi(n) = (\exp(2\pi in/2), \exp(2\pi in/3), \exp(2\pi in/4), \exp(2\pi in/5), \dots)$$

Show that the map φ is injective and that the pullback topology on $\varphi(\mathbb{Z})$ coincides with its profinite topology.

EXERCISE 1.10.7. Prove that $\mathbb{R}$ (the real line) and $\mathbb{R}^2$ (the plane with the standard topology) are not homeomorphic.

Hint: Use the notion of connected set.

EXERCISE 1.10.8. Prove that the interior of any convex polygon in $\mathbb{R}^2$ is homeomorphic to $\mathbb{R}^2$.

EXERCISE 1.10.9. A topological space $(X, \mathcal{T})$ is called *regular* (or $(T3)$- space) if for any closed set $F \subset X$ and any point $x \in X \setminus F$ there exist disjoint open sets U and V such that $F \subset U$ and $x \in V$. Give an example of a Hausdorff topological space which is not regular.

EXERCISE 1.10.10. Give an example of a regular topological space which is not normal.

EXERCISE 1.10.11. Prove that any open convex subset of $\mathbb{R}^2$ is homeomorphic to $\mathbb{R}^2$.

EXERCISE 1.10.12. Prove that any compact topological space is sequentially compact.

EXERCISE 1.10.13. Prove that any sequentially compact topological space with countable base is compact.

EXERCISE 1.10.14. A point x in a topological space is called *isolated* if the one-point set $\{x\}$ is open. Prove that any compact separable Hausdorff space without isolated points contains a closed subset homeomorphic to the Cantor set.

EXERCISE 1.10.15. Find all different topologies (up to homeomorphism) on a set consisting of 4 elements which make it a connected topological space.

EXERCISE 1.10.16. Prove that the intersection of a nested sequence of compact connected subsets of a topological space is connected.

EXERCISE 1.10.17. Give an example of the intersection of a nested sequence of compact path connected subsets of a Hausdorff topological space which is not path connected.

EXERCISE 1.10.18. Let $A \subset \mathbb{R}^2$ be the set of all vectors (x, y) such that $x + y$ is a rational number and $x - y$ is an irrational number. Show that $\mathbb{R}^2 \setminus A$ is path connected.

EXERCISE 1.10.19. Prove that any compact one–dimensional manifold is homeomorphic to the circle.

EXERCISE 1.10.20. Let $f : \mathbb{S}^1 \to \mathbb{R}^2$ be a continuous map for which there are two points $a, b \in \mathbb{S}^1$ such that $f(a) = f(b)$ and f is injective on $\mathbb{S}^1 \setminus \{a\}$. Prove that $\mathbb{R}^2 \setminus f(\mathbb{S}^1)$ has exactly three connected components.

EXERCISE 1.10.21. Consider the one–parameter group of homeomorphisms of the real line generated by the map $x \to 2x$. Consider three separation properties: (T2), (T1), and

(T0) For any two points there exists an open set which contains one of them but not the other (but which one is not given in advance).

Which of these properties does the quotient topology possess?

EXERCISE 1.10.22. Consider the group $SL(2, \mathbb{R})$ of all 2×2 matrices with determinant one with the topolology induced from the natural coordinate embedding into $\mathbb{R}^4$. Prove that it is a topological group.

CHAPTER 2

ELEMENTARY HOMOTOPY THEORY

Homotopy theory, which is the main part of algebraic topology, studies topological objects up to homotopy equivalence. Homotopy equivalence is a weaker relation than topological equivalence, i.e., homotopy classes of spaces are larger than homeomorphism classes. Even though the ultimate goal of topology is to classify various classes of topological spaces up to a homeomorphism, in algebraic topology, homotopy equivalence plays a more important role than homeomorphism, essentially because the basic tools of algebraic topology (homology and homotopy groups) are invariant with respect to homotopy equivalence, and do not distinguish topologically nonequivalent, but homotopic objects.

The first examples of homotopy invariants will appear in this chapter: degree of circle maps in Section 2.4, the fundamental group in Section 2.8 and higher homotopy groups in Section 2.10, while homology groups will appear and will be studied later, in Chapter 8. In the present chapter, we will see how effectively homotopy invariants work in simple (mainly low-dimensional) situations.

2.1. Homotopy and homotopy equivalence

2.1.1. Homotopy of maps. It is interesting to point out that in order to define the homotopy equivalence, a relation between *spaces*, we first need to consider a certain relation between *maps*, although one might think that spaces are more basic objects than maps between spaces.

DEFINITION 2.1.1. Two continuous maps $f_0, f_1 : X \to Y$ between topological spaces are said to be *homotopic* if there exists a a continuous map F : $X \times [0,1] \to Y$ (the *homotopy*) that F *joins* f_0 *to* f_1, i.e., if we have $F(i, \cdot) = f_i$ for $i = 1, 2$.

A map $f : X \to Y$ is called *null-homotopic* if it is homotopic to a constant map $c : X \to \{y_0\} \subset Y$. If $f_0, f_1 : X \to Y$ are homeomorphisms, they are called *isotopic* if they can be joined by a homotopy F (the *isotopy*) which is a homeomorphism $F(t, \cdot)$ for every $t \in [0,1]$.

If two maps $f, g : X \to Y$ are homotopic, we write $f \simeq g$.

maybe add a p

EXAMPLE 2.1.2. The identity map id: $\mathbb{D}^2 \to \mathbb{D}^2$ and the constant map c_0 : $\mathbb{D}^2 \to 0 \in \mathbb{D}^2$ of the disk $\mathbb{D}^2$ are homotopic. A homotopy between them may be defined by $F(t, (\rho, \varphi)) = ((1-t) \cdot \rho, \varphi)$, where (ρ, φ) are polar coordinates in $\mathbb{D}^2$. Thus the identity map of the disk is null homotopic.

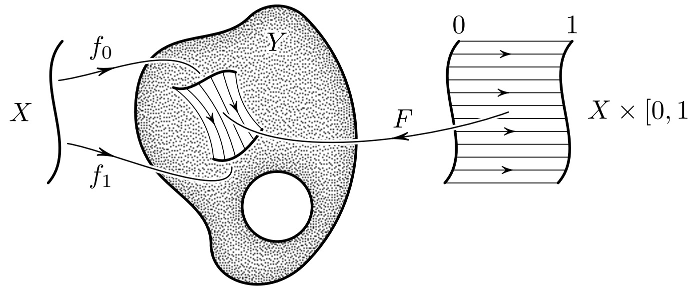

FIGURE 2.1.1. Homotopic maps

EXAMPLE 2.1.3. If the maps $f, g : X \to Y$ are both null-homotopic and Y is path connected, then they are homotopic to each other.

Indeed, suppose a homotopy F joins f with the constant map to the point $a \in Y$, and a homotopy G joins g with the constant map to the point $b \in Y$. Let $c : [0,1] \to Y$ be a path from a to b. Then the following homotopy

$$H(t,x) := \begin{cases} F(x, 3t) & \text{when } 0 \le t \le \frac{1}{3}, \\ c(3t-1) & \text{when } \frac{1}{3} \le t \le \frac{2}{3} \\ G(x, 3-3t) & \text{when } \frac{2}{3} \le t \le 1. \end{cases}$$

joins the map f to g.

EXAMPLE 2.1.4. If A is the annulus $A = \{(x,y)|1 \le x^2 + y^2 \le 2\}$, and the circle $\mathbb{S}^1 = \{z \in \mathbb{C} : |z| = 1\}$ is mapped homeomorphically to the outer and inner boundary circles of A according to the rules $f : e^{i\varphi} \mapsto (2, \varphi)$ and $g : e^{i\varphi} \mapsto (1, \varphi)$ (here we are using the polar coordinates (r, φ)) in the $(x,y) - plane$), then f and g are homotopic.

Indeed, $H(t, \varphi) := (t+1, \varphi)$ provides the required homotopy.

Further, it should be intuitively clear that neither of the two maps f or g is null homotopic, but at this point we do not possess the appropriate techniques for proving that fact.

2.1.2. Homotopy equivalence. To motivate the definition of homotopy equivalent spaces let us write the definition of homeomorphic spaces in the following form: topological spaces X and Y are homeomorphic if there exist maps $f : X \to Y$ and $g : Y \to X$ such that

$$f \circ g = \mathrm{Id}_X \quad \text{and} \quad g \circ f = \mathrm{Id}_Y \,.$$

If we now replace equality by homotopy we obtain the desired notion:

DEFINITION 2.1.5. Two topological spaces X, Y are called *homotopy equivalent* if there exist maps $f : X \to Y$ and $g : Y \to X$ such that

$$f \circ g : X \to X \quad \text{and} \quad g \circ f : Y \to Y$$

are homotopic to the corresponding identities Id_X and Id_Y.

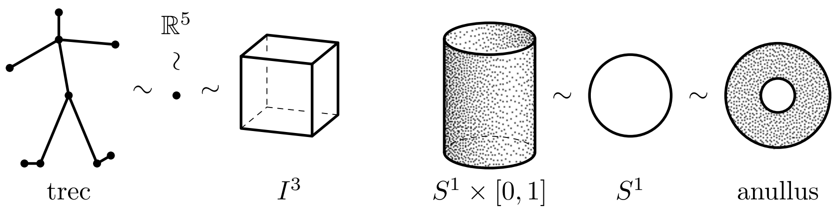

FIGURE 2.1.2. Homotopy equivalent spaces

EXAMPLE 2.1.6. The point, the disk, the Euclidean plane are all homotopy equivalent. To show that pt$\simeq$ $\mathbb{R}^2$, consider the maps f : pt $\to 0 \in \mathbb{R}^2$ and $g : \mathbb{R}^2 \to$pt. Then $g \circ f$ is just the identity of the one point set pt, while the map $f \circ g : \mathbb{R}^2 \to \mathbb{R}^2$ is joined to the identity of $\mathbb{R}^2$ by the homotopy $H(t,(r,\varphi)) := ((1-t)r,\varphi)$.

EXAMPLE 2.1.7. The circle and the annulus are homotopy equivalent. Mapping the circle isometrically on the inner boundary of the annulus and projecting the entire annulus along its radii onto the inner boundary, we obtain two maps that comply with the definition of homotopy equivalence.

PROPOSITION 2.1.8. *The relation of being homotopic* (*maps*) *and being homotopy equivalent* (*spaces*) *are equivalence relations in the technical sense, i.e., are reflexive, symmetric, and transitive.*

PROOF. The proof is quite straightforward. First let us check transitivity for maps and reflexivity for spaces.

Suppose $f \simeq g \simeq h$. Let us prove that $f \simeq h$. Denote by F and G the homotopies joining f to g and g to h, respectively. Then the homotopy

$$H(t,x) := \begin{cases} F(2t) & \text{when } t \le \frac{1}{2}, \\ G(2t-1) & \text{when } t \ge \frac{1}{2} \end{cases}$$

joins f to h.

Now let us prove that for spaces the relation of homotopy equivalence is reflexive, i.e., show that for any topological space X we have $X \simeq X$. But the pair of maps $(\mathrm{id}_X, \mathrm{id}_X)$ and the homotopy given by $H(t,x) := x$ for any t shows that X is indeed homotopy equivalent to itself.

The proofs of the other properties are similar and are omitted. □

maybe add p

PROPOSITION 2.1.9. *Homeomorphic spaces are homotopy equivalent.*

PROOF. If $h : X \to Y$ is a homeomorphism, then $h \circ h^{-1}$ and $h^{-1} \circ h$ are the identities of Y and X, respectively, so that the homotopy equivalence of X and Y is an immediate consequence of the reflexivity of that relation. □

In our study of topological spaces in the previous chapter, the main equivalence relation was homeomorphism. In homotopy theory, its role is played by homotopy equivalence. As we have seen, homeomorphic spaces are homotopy equivalent. The converse is not true, as simple examples show.

the picture above

EXAMPLE 2.1.10. Euclidean space $\mathbb{R}^n$ and the point are homotopy equivalent but not homeomorphic since there is no bijection between them. Open and closed interval are homotopy equivalent since both are homotopy equivalent to a point but not homeomorphic since closed interval is compact and open is not.

EXAMPLE 2.1.11. The following five topological spaces are all homotopy equivalent but any two of them are not homeomorphic:

- the circle $\mathbb{S}^1$,
- the open cylinder $\mathbb{S}^1 \times \mathbb{R}$,
- the annulus $A = \{(x, y)|1 \leq x^2 + y^2 \leq 2\}$,
- the solid torus $\mathbb{S}^1 \times \mathbb{D}^2$,
- the Möbius strip.

In all cases one can naturally embed the circle into the space and then project the space onto the embedded circle by gradually contracting remaining directions. Proposition 2.2.8 below also works for all cases but the last.

Absence of homeomorphisms is shown as follows: the circle becomes disconnected when two points are removed, while the other spaces are not; the annulus and the solid torus are compact, the open cylinder and the Möbius strip are not. The remaining two pairs are a bit more tricky since thy require making intuitively obvious statement rigorous: (i) the annulus has two boundary components and the solid torus one, and (ii) the cylinder becomes disconnected after removing any subset homeomorphic to the circle[1] while the Möbius strip remains connected after removing the middle circle.

As is the case with homeomorphisms in order to establish that two spaces are homotopy equivalent one needs just to produce corresponding maps while in order to establish the absence of homotopy equivalence an invariant is needed which can be calculated and shown to be different for spaces in question. Since homotopy equivalence is a more robust equivalence relation that homeomorphism there are fewer invariants and many simple homeomorphism invariants do not work, e.g. compactness and its derivative connectedness after removing one or more points and so on. In particular, we still lack means to show that the spaces from two previous examples are not homotopy equivalent. Those means will be provided in Section 2.4

2.2. Contractible spaces

Now we will study properties of contractible spaces, which are, in a natural sense, the trivial objects from the point of view of homotopy theory.

[1]This follows from the Jordan Curve Theorem Theorem 5.1.2

2.2.1. Definition and examples. As we will see from the definition and examples, contractible spaces are connected topological objects which have no "holes", "cycles", "apertures" and the like.

DEFINITION 2.2.1. A topological space X is called *contractible* if it is homotopically equivalent to a point. Equivalently, a space is contractible if its identity map is null-homotopic.

EXAMPLE 2.2.2. Euclidean and complex spaces $\mathbb{R}^n$, $\mathbb{C}^n$ are contractible for all n. So is the closed n-dimensional ball (disc) $\mathbb{D}^n$, any tree (graph without cycles; see Section 2.3), the wedge of two disks. This can be easily proven by constructing homotopy equivalence On the other hand, the sphere $\mathbb{S}^n$, $n \geq 0$, the torus $\mathbb{T}^n$, any graph with cycles or multiple edges are all not contractible. To prove this one needs to construct some*invariants*, i.e. quantities which are equal for homotopy equivalent spaces. An this point we do not have such invariants yet.

PROPOSITION 2.2.3. *Any convex subset of* $\mathbb{R}^n$ *is contractible.*

PROOF. Let C be a convex set in $\mathbb{R}^n$ ant let $x_0 \in C$, define

$$h(x,t) = x_0 + (1-t)(x - x_0).$$

By convexity for any $t \in [0,1]$ we obtain a map of C into itself. This is a homotopy between the identity and the constant map to x_0 □

REMARK 2.2.4. The same proof works for a broader class of sets than convex, namely star-shaped. A set $S \subset \mathbb{R}^n$ is called *star-shaped* if there exists a point x_0 such that the intersection of any half line with endpoint x_0 with S is an interval. hence any star-shaped set is contractible.

2.2.2. Properties. Contractible spaces have nice intrinsic properties and also behave well under maps.

PROPOSITION 2.2.5. *Any contractible space is path connected.*

PROOF. Let $x_1, x_2 \in X$, where X is contractible. Take a homotopy h between the identity and a constant map, to, say x_0. Let

$$f(t) := \begin{cases} h(x, 2t) & \text{when } t \leq \frac{1}{2}, \\ h(y, 2t-1) & \text{when } t \geq \frac{1}{2}. \end{cases}$$

Thus f is a continuous map of $[0,1]$ to X with $f(0) = x$ and $f(1) = y$. □

PROPOSITION 2.2.6. *If the space* X *is contractible, then any map of this space* $f : X \to Y$ *is null homotopic.*

PROOF. By composing the homotopy taking X to a point p and the map f, we obtain a homotopy of f and the constant map to $f(p)$. □

PROPOSITION 2.2.7. *If the space Y is contractible, then any map to this space $f : X \to Y$ is null homotopic.*

PROOF. By composing the map f with the homotopy taking Y to a point and, we obtain a homotopy of f and the constant map to that point. $\square$

PROPOSITION 2.2.8. *If X is contractible, then for any topological space Y the product $X \times Y$ is homotopy equivalent to X.*

PROOF. If $h : Y \times [0,1] \to Y$ is a homotopy between the identity and a constant map of Y,that is, $h(y,0) = y$ and $h(y,1) = y_0$. Then for the map $H := \mathrm{Id}_X \times h$ one has $H(x,y,0) = (x,y)$ and $H(x,y,1) = (x,y_0)$. Thus the projection $\pi_1 : (x,y) \mapsto x$ and the embedding $i_{y_0} : x \mapsto (x,y_0)$ provide a homotopy equivalence. $\square$

2.3. Graphs

In the previous section, we discussed contractible spaces, the simplest topological spaces from the homotopy point of view, i.e., those that are homotopy equivalent to a point. In this section, we consider the simplest type of space from the point of view of dimension and local structure: graphs, which may be described as one-dimensional topological spaces consisting of line segments with some endpoints identified.

We will give a homotopy classification of graphs, find out what graphs can be embedded in the plane, and discuss one of their homotopy invariants, the famous Euler characteristic.

2.3.1. Main definitions and examples. Here we introduce (nonoriented) graphs as classes of topological spaces with an edge and vertex structure and define the basic related notions, but also look at abstract graphs as very general combinatorial objects. In that setting an extra orientation structure becomes natural.

DEFINITION 2.3.1. A (nonoriented) *graph* G is a topological space obtained by taking a finite set of line segments (called *edges* or *links*) and identifying some of their endpoints (called *vertices* or *nodes*).

Thus the graph G can be thought of as a finite sets of points (vertices) some of which are joined by line segments (edges); the sets of vertices and edges are denoted by $V(G)$ and $E(G)$, respectively. If a vertex belongs to an edge, we say that the vertex is *incident* to the edge or the edge is incident to the vertex. A *morphism* of graphs is a map of vertices and edges preserving incidence, an *isomorphism* is a bijective morphism.

It the two endpoints of an edge are identified, such an edge is called a *loop*. A *path* (or *chain*) is a ordered set of edges such that an endpoint of the first edge coincides with an endpoint of the second one, the other endpoint of the second edge coincides with an endpoint of the third edge, and so on, and finally an endpoint of the last edge coincides with an endpoint of the previous one. A closed path (i.e.,

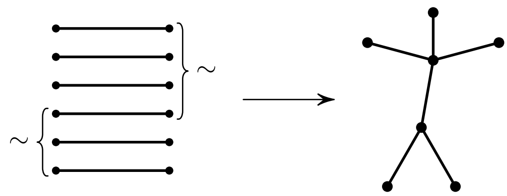

FIGURE 2.3.1. Constructing a graph by identifying endpoints of segments

a path whose first vertex coincides with its last one) is said to be a *cycle*; a loop is regarded as a particular case of a cycle.

A *tree* is a graph without cycles.

A graph is called *connected* if any two vertices can be joined by a path. This is equivalent to the graph being connected (or path-connected) as a topological space.

The number of edges with endpoints at a given vertex is called the *degree* of this vertex, the *degree of a graph* is the maximal degree of all its vertices.

A *complete graph* is a graph such that each pair of distinct vertices is joined by exactly one edge. check the

EXERCISE 2.3.1. Prove that any graph can be embedded into $\mathbb{R}^3$, i.e. it is isomorphic to a graph which is a subset of the three-dimensional space $\mathbb{R}^3$.

A graph is called *planar* if it is isomorphic to a graph which is a subset of the plane $\mathbb{R}^2$.

EXAMPLE 2.3.2. The sets of vertices and edges of the n-simplex constitute a graph, which is connected and complete, and whose vertices are all of degree $n+1$. The sets of edges of an n-dimensional cube constitute a connected graph whose vertices are all of degree n, but which is not complete (if $n \geq 2$).

EXAMPLE 2.3.3. The figure shows two important graphs, $K_{3,3}$ and K_5, both of which are nonplanar. The first is the formalization of a famous (unsolvable) problem: to find paths joining each of three houses to each of three wells so that the paths never cross. In practice would have to build bridges or tunnels. The second is the complete graph on five vertices. The proof of their nonplanarity will be discussed on the next subsection.

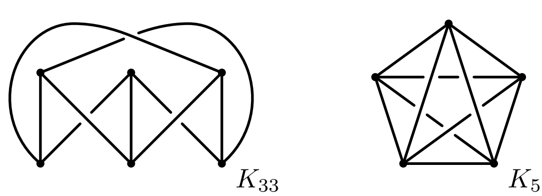

FIGURE 2.3.2. Two nonplanar graphs: $K_{3,3}$ and K_5

DEFINITION 2.3.4. An *oriented graph* is a graph with a chosen direction on each edge. Paths and cycles are defined as above, except that the edges must be

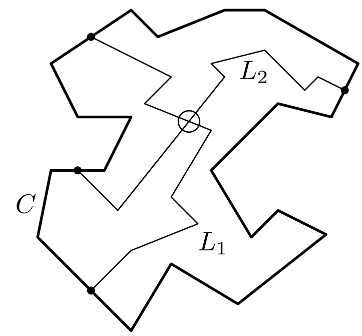

FIGURE 2.3.3. The polygonal lines L_1 and L_2 must intersect

coherently oriented. Vertices with only one edge are called *roots* if the edge is oriented away from the vertex, and *leaves* if it is oriented towards the vertex.

2.3.2. Planarity of graphs. The goal of this subsection is to prove that the graph $K_{3,3}$ is nonplanar, i.e., possesses no topological embedding into the plane $\mathbb{R}^2$. To do this, we first prove the polygonal version of the Jordan curve theorem and show that the graph $K_{3,3}$ has no polygonal embedding into the plane, and then show that it has no topological embedding in the plane.

PROPOSITION 2.3.5. [The Jordan curve theorem for broken lines] *Any broken line C in the plane without self-intersections splits the plane into two path connected components and is the boundary of each of them.*

PROOF. Let D be a small disk which C intersects along a line segment, and thus divides D into two (path) connected components. Let p be any point in $\mathbb{R}^2 \setminus C$. From p we can move along a polygonal line as close as we like to C and then, staying close to C, move inside D. We will then be in one of the two components of $D \setminus C$, which shows that $\mathbb{R}^2 \setminus C$ has *no more than two* components.

It remains to show that $\mathbb{R}^2 \setminus C$ is not path connected. Let ρ be a ray originating at the point $p \in \mathbb{R}^2 \setminus C$. The ray intersects C in a finite number of segments and isolated points. To each such point (or segment) assign the number 1 if C crosses ρ there and 0 if it stays on the same side. Consider the parity $\pi(p)$ of the sum S of all the assigned numbers: it changes continuously as ρ rotates and, being an integer, $\pi(p)$ is constant. Clearly, $\pi(p)$ does not change inside a connected component of $\mathbb{R}^2 \setminus C$. But if we take a segment intersecting C at a non-zero angle, then the parity π at its end points differs. This contradiction proves the proposition. $\square$

We will call a closed broken line without self-intersections a *simple polygonal line*.

COROLLARY 2.3.6. *If two broken lines L_1 and L_2 without self-intersections lie in the same component of $\mathbb{R}^2 \setminus C$, where C is a simple closed polygonal line, with their endpoints on C in alternating order, then L_1 and L_2 intersect.*

PROOF. The endpoints a and c of L_1 divide the polygonal curve C into two polygonal arcs C_1 and C_2. The curve C and the line L_1 divide the plane into three path connected domains: one bounded by C, the other two bounded by the closed curves $C_i \cup L$, $i = 1, 2$ (this follows from Proposition 2.3.5). Choose points b and d on L_2 close to its endpoints. Then b and d must lie in different domains bounded by L_1 and C and any path joining them and not intersecting C, in particular L_2, must intersect L_1. □

PROPOSITION 2.3.7. *The graph $K_{3,3}$ cannot be polygonally embedded in the plane.*

PROOF. Let us number the vertices $x_1, \dots, x_6$ of $K_{3,3}$ so that its edges constitute a closed curve $C := x_1x_2x_3x_4x_5x_6$, the other edges being

$$E_1 := x_1x_4, \quad E_2 := x_2x_5, \quad E_3 := x_3x_6.$$

Then, if $K_{3,3}$ lies in the plane, it follows from Proposition 2.3.5 that C divides the plane into two components. One of the two components must contain at least two of the edges E_1, E_2, E_3, which then have to intersect (by Corollary 2.3.6). This is a contradiction which proves the proposition. □

THEOREM 2.3.8. *The graph $K_{3,3}$ is nonplanar, i.e., there is no topological embedding $h : K_{3,3} \hookrightarrow \mathbb{R}^2$.*

The theorem is an immediate consequence of the nonexistence of a PL-embedding of $K_{3,3}$ (Proposition 2.3.7) and the following lemma.

LEMMA 2.3.9. *If a graph G is planar, then there exists a polygonal embedding of G into the plane.*

PROOF. Given a graph $G \subset \mathbb{R}^2$, we first modify it in small disk neighborhoods of the vertices so that the intersection of (the modified graph) G with each disk is the union of a finite number of radii of this disk. Then, for each edge, we cover its complement to the vertex disks by disks disjoint from the other edges, choose a finite subcovering (by compactness) and, using the chosen disks, replace the edge by a polygonal line. □

We conclude this subsection with a beautiful theorem, which gives a simple geometrical obstruction to the planarity of graphs. We do not present the proof (which is not easy), because this theorem, unlike the previous one, is not used in the sequel.

small print for pa
the main line: no

THEOREM 2.3.10. [Kuratowski] *A graph is nonplanar if and only if it contains, as a topological subspace, the graph $K_{3,3}$ or the graph K_5.*

REMARK 2.3.11. The words “as a topological subspace” are essential in this theorem. They cannot be replaced by “as a subgraph”: if we subdivide an edge of K_5 by adding a vertex at its midpoint, then we obtain a nonplanar graph that does *not* contain either $K_{3,3}$ or K_5.

EXERCISE 2.3.2. Can the graph $K_{3,3}$ be embedded in (a) the Möbius strip, (b) the torus?

EXERCISE 2.3.3. Is there a graph that cannot be embedded into the torus?

EXERCISE 2.3.4. Is there a graph that cannot be embedded into the Möbius strip?

2.3.3. Euler characteristic of graphs and plane graphs. The *Euler characteristic* of a graph G is defined as

$$\chi(G) \ := \ V - E,$$

where V is the number of vertices and E is the number of edges.

The *Euler characteristic* of a graph G without loops *embedded in the plane* is defined as

$$\chi(G) \ := \ V - E + F,$$

where V is the number of vertices and E is the number of edges of G, while F is the number of connected components of $\mathbb{R}^2 \setminus G$ (including the unbounded component).

THEOREM 2.3.12. [Euler Theorem] *For any connected graph G without loops embedded in the plane,* $\chi(G) \ = \ 2.$

PROOF. At the moment we are only able to prove this theorem for polygonal graphs. For the general case we will need Jordan curve Theorem Theorem 5.1.2. The proof will be by induction on the number of edges. Without loss of generality, we can assume (by Lemma 2.3.9) that the graph is polygonal. For the graph with zero edges, we have $V = 1$, $E = 0$, $F = 1$, and the formula holds. Suppose it holds for all graphs with n edges; then it is valid for any connected subgraph H of G with n edges; take an edge e from G which is not in H but incident to H, and add it to H. Two cases are possible.

Case 1. Only one endpoint of e belongs to H. Then F is the same for G as for H and both V and E increase by one.

Case 2. Both endpoints of e belong to to H. Then e lies inside a face of H and divides it into two.[2] Thus by adding e we increase both E and F by one and leave V unchanged. Hence the Euler characteristic does not change. □

2.3.4. Homotopy classification of graphs. It turns out that, from the viewpoint of homotopy, graphs are classified by their Euler characteristic (which is therefore a *complete* homotopy invariant.)

EXERCISE 2.3.5. Prove that any tree is homotopy equivalent to a point.

THEOREM 2.3.13. *Any connected graph G is homotopy equivalent to the wedge of k circles, with* $k = \chi(G) - 1$.

[2] It is here that we need the conclusion of Jordan curve Theorem Theorem 5.1.2 in the case of general graphs. The rest of the argument remains the same as for polygonal graphs.

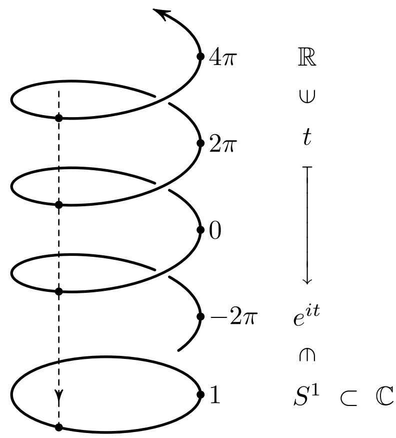

FIGURE 2.4.1. Exponential map

PROOF. Consider a maximal tree T which is a subgraph of G. The graph W obtained by identifying T into a single vertex p is homotopically equivalent to G. But any edge of W whose one endpoint is p must be a loop since otherwise T would not be a maximal tree in G. Since W is connected it has a single vertex p and hence is a wedge of several loops. □

At this point we do not know yet that wedges of different numbers of circles are mutually not homotopically equivalent or, for that matter that they are not contractible. This will be shown with the use of the first non-trivial homotopy invariant which we will study in the next section. This will of course also imply that the Euler characteristic of a graph is invariant under homotopy equivalence.

2.4. Degree of circle maps

Now we will introduce a homotopy invariant for maps of the circle to itself. It turns out that this invariant can easily be calculated and have many impressive applications. Some of those applications are presented in three subsequent sections.

2.4.1. The exponential map. Recall the relation between the circle $\mathbb{S}^1 = \mathbb{R}/\mathbb{Z}$ and the line $\mathbb{R}$. There is a projection $\pi\colon \mathbb{R} \to \mathbb{S}^1$, $x \mapsto [x]$, where $[x]$ is the equivalence class of x in $\mathbb{R}/\mathbb{Z}$. Here the integer part of a number is written $\lfloor\cdot\rfloor$ and $\{\cdot\}$ stands for the fractional part.

PROPOSITION 2.4.1. *If $f\colon \mathbb{S}^1 \to \mathbb{S}^1$ is continuous, then there exists a continuous map $F\colon \mathbb{R} \to \mathbb{R}$, called a* lift *of f to $\mathbb{R}$, such that*

$$f \circ \pi = \pi \circ F, \tag{2.4.1}$$

that is, $f([z]) = [F(z)]$. Such a lift is unique up to an additive integer constant and $\deg(f) := F(x+1) - F(x)$ *is an integer independent of $x \in \mathbb{R}$ and the lift F. It is called the* degree *of f. If f is a homeomorphism, then $|\deg(f)| = \pm 1$.*

PROOF. Existence: Pick a point $p \in \mathbb{S}^1$. Then we have $p = [x_0]$ for some $x_0 \in \mathbb{R}$ and $f(p) = [y_0]$ for some $y_0 \in \mathbb{R}$. From these choices of x_0 and y_0 define $F\colon \mathbb{R} \to \mathbb{R}$ by requiring that $F(x_0) = y_0$, that F be continuous, and that $f([z]) = [F(z)]$ for all $z \in \mathbb{R}$. One can construct such an F, roughly speaking, by varying the initial point p continuously, which causes $f(p)$ to vary continuously. Then there is no ambiguity of how to vary x and y continuously and thus $F(x) = y$ defines a continuous map.

To elaborate, take a $\delta > 0$ such that

$$d([x],[x']) \leq \delta \text{ implies } d(f([x]), f([x'])) < 1/2.$$

Then we can define F on $[x_0 - \delta, x_0 + \delta]$ as follows: If $|x - x_0| \leq \delta$ then $d(f([x]), q) < 1/2$ and there is a unique $y \in (y_0 - 1/2, y_0 + 1/2)$ such that $[y] = f([x])$. Define $F(x) = y$. Analogous steps extend the domain by another δ at a time, until F is defined on an interval of unit length. (One needs to check consistency, but it is straightforward.) Then $f([z]) = [F(z)]$ defines F on $\mathbb{R}$.

Uniqueness: Suppose $\tilde{F}$ is another lift. Then $[\tilde{F}(x)] = f([x]) = [F(x)]$ for all x, meaning $\tilde{F} - F$ is always an integer. But this function is continuous, so it must be constant.

Degree: $F(x+1) - F(x)$ is an integer (now evidently independent of the choice of lift) because

$$[F(x+1)] = f([x+1]) = f([x]) = [F(x)].$$

By continuity $F(x+1) - F(x) =: \deg(f)$ must be a constant.

Invertibility: If $\deg(f) = 0$, then $F(x+1) = F(x)$ and thus F is not monotone. Then f is noninvertible because it cannot be monotone. If $|\deg(f)| > 1$, then $|F(x+1) - F(x)| > 1$ and by the Intermediate Value Theorem there exists a $y \in (x, x+1)$ with $|F(y) - F(x)| = 1$, hence $f([y]) = f([x])$, and $[y] \neq [x]$, so f is noninvertible. □

2.4.2. Homotopy invariance of the degree. Here we show that the degree of circle maps is a homotopy invariant and obtain some immediate corollaries of this fact.

PROPOSITION 2.4.2. *Degree is a homotopy invariant.*

PROOF. The lift construction can be simultaneously applied to a continuous family of circle maps to produce a continuous family of lifts. Hence the degree must change continuously under homotopy. Since it is an integer, it is in fact constant. □

COROLLARY 2.4.3. *The circle is not contractible.*

PROOF. The degrees of any constant map is zero, whereas for the identity map it is equal to one. □

THEOREM 2.4.4. *Degree is a complete homotopy invariant of circle self–maps: for any $m \in \mathbb{Z}$ any map of degree m is homotopic to the map*

$$E_m := x \mapsto mx \pmod 1.$$

PROOF. Obviously, the map E_m lifts to the linear map $x \mapsto mx$ of R. On the other hand, every lift F of a degree m map f has the form $F(x) = mx + H(x)$, where H is a periodic function with period one. Thus the family of maps

$$F_t(x) := mx + (1-t)H(x)$$

are lifts of a continuous family of maps of S^1 which provide a homotopy between f and E_m. □

Since $E_m \circ E_n = E_{mn}$ we obtain

COROLLARY 2.4.5. *Degree of the composition of two maps is equal to the product of their degrees.*

EXERCISE 2.4.1. Show that any continuous map $f : \mathbb{S}^1 \to \mathbb{S}^1$ has at least $|\deg f - 1|$ fixed points.

EXERCISE 2.4.2. Prove Corollary 2.4.5 directly, not using Theorem 2.4.4.

EXERCISE 2.4.3. Given the maps $f : \mathbb{S}^1 \to \mathbb{D}^2$ and $g : \mathbb{D}^2 \to \mathbb{S}^1$, what can be said about the degree of their composition.

2.4.3. Degree and wedges of circles. In order to complete homotopy classification of graphs started in Section 2.3.4 we need to proof the following fact which will be deduced from the degree theory for circle maps.

PROPOSITION 2.4.6. *The wedges of k circles for $k = 0, 1, 2, \ldots$ are pairwise not homotopy equivalent.*

PROOF. We first show that the wedge of any number of circles is not contractible. For one circle this has been proved already (2.4.3). Let W be the wedge of $k > 1$ circles and $p \in W$ be the common point of the circles. If W is contractible then the identity map Id_W of W is homotopic to the constant map c_P of W to p. Let S be one of the circles comprising W and let U be the union of remaining circles. Then one can identify U into a single point (naturally identified with p and project the homotopy to the identification space which is naturally identified with the circle S and thus provides a homotopy between the identity and a constant map on the circle, a contradiction. More specifically we apply the following process which looks like cutting the graph of a continuous function at a constant level when the function exceeds this level. As long as the images of a point $x \in S$ stay in S we change nothing. When it reaches p and leaves S we replace the images by the constant p.

Now assume that the wedge W of m circles is homotopically equivalent to the wedge of $n < m$ circles which can be naturally identified with a subset U of W consisting of n circles. This implies that there exists a homotopy between Id_W and the map $c_U : W \to W$ which is equal to the identity on U and maps $m - n$ circles comprising $W \setminus U$ into the common point p of all circles in W. As before, we identify U into a point and project the homotopy into the identification space which is naturally identified with the wedge of $m - n$ circles. Thus we obtain a homotopy between a homotopy between the identity map and the constant map c_p which is impossible by the previous argument. $\square$

es not follow from
uivalence directly.
argued or replaces
another argument

Now we can state the homotopy classification of graphs as follows.

COROLLARY 2.4.7. *Two graphs are homotopy equivalent if and only if they have the same Euler characteristic. Any graph with Euler characteristic E is homotopy equivalent to the wedge of $E + 1$ circles.*

2.4.4. Local definition of degree. One of the central ideas in algebraic topology is extension of the notion of degree of a self-map from circles to spheres of arbitrary dimension and then to a broad class of compact manifold. Definition which follows from Proposition 2.4.1 stands no chance of generalization since the exponential map is a phenomenon specific for the circle and, for example in has no counterparts for spheres of higher dimensions. Now we give another definition which is equivalent to the previous one for the circle but can be generalized to other manifolds.

We begin with piecewise strictly monotone maps of the circle into itself. For such a map every point $x \in \mathbb{S}^1$ has finitely many pre-images and for if we exclude finitely many values at the endpoints of the interval of monotonicity each pre-image $y \in f^{-1}(x)$ lies on a certain interval of monotonicity where the function f either "increases", i.e. preserves orientation on the circle or "decreases", i.e. reverses orientation. In the first case we assign number 1 to the point y and call it a positive pre-image and in the second the number -1 and call it a negative pre-image of x. Adding those numbers for all $y \in f^{-1}(x)$ we obtain an integer which we denote $d(x)$.

provide picture(s)

THEOREM 2.4.8. *The number $d(x)$ is independent of x and is equal to the degree of f.*

REMARK 2.4.9. Since any continuous map of the circle can be arbitrary well approximated by a piecewise monotone map (in fact, even by a piecewise linear one) and by the above theorem the number thus defined for piecewise monotone maps (call it the local degree) is the same for any two sufficiently close maps we can *define* degree of an arbitrary continuous map $f\colon \mathbb{S}^1 \to \mathbb{S}^1$ as the the local degree of any piecewise monotone map g sufficiently close to f. This is a "baby version" of the procedure which will be developed for other manifolds in **??**.

PROOF. Call a value $x \in \mathbb{S}^1$ *critical* if $x = f(y)$ where x is an endpoint of an interval of monotonicity for f which we will call critical points. Obviously $d(x)$

does not change in a neighborhood of any non-critical value. It does not change at a critical value either since each critical value is the image of several critical points and near each such point either there is one positive and one negative pre-image for nearby values on one side and none on the other or vise versa. Thus $d(x)$ is constant which depends only on the map f and can thus be denoted by $d(f)$.

For any piecewise monotone map f let as call its *piecewise linear approximation* f_{PL} the map which has the same intervals of monotonicity and is linear on any of them. Obviously $d(f_{PL}) = d(f)$; this follows from a simple application of the intermediate value theorem from calculus. Consider the straight-line deformation of the map f_{PL} to the linear map $E_{\deg f}$. Notice that since f_{PL} is homotopic to f (by the straight line on each monotonicity interval) $\deg f_{PL} = \deg f$. This homotopy passes through piecewise linear maps which we denote by g_t and hence the local degree is defined. A small point is that for some values of t the map g_t may be constant on certain intervals of monotonicity of of f but local degree is defined for such maps as well. It remains to notice that the local degree does not change during this deformation. But this is obvious since any non-critical value of g_t remains non-critical with a small change of t and for each t all but finitely many values are non-critical. Since local degree can be calculated at any non-critical value this shows that

$$d(f) = d(f_{PL}) = d(E_{\deg f}) = \deg f.$$

□

2.5. Brouwer fixed point theorem in dimension two

In the general case, the Brouwer theorem says that any (continuous) self-map of the disk $\mathbb{D}^n$ (a closed ball in $\mathbb{R}^n$) has a fixed point, i.e., there exists a $p \in \mathbb{D}^n$ such that $f(p) = p$.

The simplest instance of this theorem (for $n = 1$) is an immediate corollary of the intermediate value theorem from calculus since a continuous map f of a closed interval $[a, b]$ into itself can be considered as a real-values function such that $f(a) \geq a$ and $f(b) \geq b$. Hence by the intermediate value theorem the function $f(x) - x$ has a zero on $[a, b]$.

The proof in dimension two is based on properties of the degree.

THEOREM 2.5.1. [Brouwer fixed-point theorem in dimension two.] *Any continuous map of a closed disk into itself (and hence of any space homeomorphic to the disk) has a fixed point.*

PROOF. We consider the standard closed disc

$$\mathbb{D}^2 := \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}.$$

Suppose $f : \mathbb{D}^2 \to \mathbb{D}^2$ is a continuous map without fixed points. For $p \in \mathbb{D}^2$ consider the open halfline (ray) beginning at $F(p)$ and passing through the point p. This halfline intersects the unit circle $\mathbb{S}^1$, which is the boundary of the disc $\mathbb{D}^2$, at a single point which we will denote by $h(p)$. Notice that for $p \in \partial\mathbb{D}^2$, $h(p) = p$

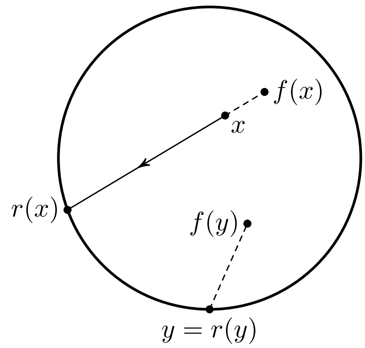

FIGURE 2.5.1. Retraction of the disk onto the circle

The map $h : \mathbb{D}^2 \to \partial\mathbb{D}^2$ thus defined is continuous by construction (exactly because f has no fixed points) and is homotopic to the identity map $\mathrm{Id}_{\mathbb{D}^2}$ via the straight-line homotopy $H(p,t) = (1-t)p + th(p)$. Now identify $\partial\mathbb{D}^2$ with the unit circle $\mathbb{S}^1$. Taking the composition of h with this identification, we obtain a map $\mathbb{D}^2 \to \mathbb{S}^1$, which we will denote by g. Let $i : \mathbb{S}^1 \to \mathbb{D}^2$ be the standard embedding. We have

$$g \circ i : \mathbb{S}^1 \to \mathbb{S}^1 = \mathrm{Id}_{\mathbb{S}^1}, \ \ i \circ g = h \ \text{ is homotopic to } \mathrm{Id}_{\mathbb{D}^2} .$$

Thus the pair (i, g) gives a homotopy equivalence between $\mathbb{S}^1$ and $\mathbb{D}^2$.

But this is impossible, since the disc is contractible and the circle is not (Corollary 2.4.3). Hence such a map h cannot be constucted; this implies that F has a fixed point at which the halfline in question cannot be uniquely defined. □

EXERCISE 2.5.1. Deduce the general form of the Brouwer fixed–point theorem: *Any continuous map of a closed n-disc into itself has a fixed point*, from the fact that the identity map on the sphere of any dimension is not null homotopic. The latter fact will be proved later (**??**).

2.6. Index of a point w.r.t. a curve

In this section we study curves and points lying in the plane $\mathbb{R}^2$ and introduce an important invariant: the index $\mathrm{ind}(p, \gamma)$ of a point p with respect to a curve $\gamma : \mathbb{S}^1 \to \mathbb{R}^2$. This invariant has many applications, in particular it will help us prove the so-called "Fundamental Theorem of Algebra" in the next section.

2.6.1. Main definition and examples. By a *curve* we mean the image $C = f(\mathbb{S}^1)$ of a continuous map $f : \mathbb{S}^1 \to \mathbb{R}^2$, not necessarily injective. Recall that C is compact by Proposition 1.5.11 Let p be a point in the open complement $\mathbb{R}^2 - C$ of the curve. The complement is nonempty since C is compact but R^2 is not. Notice however that C may have an interior if f is a so-called *Peano curve* **??** or somehting similar. Denote by φ the angular parameter on $\mathbb{S}^1$ and by V_φ the vector joining the points p and $f(\varphi)$. As φ varies from 0 to 2π, the endpoint of the unit vector $V_\varphi/|V_\varphi|$ moves along the unit circle S_0 centered at p, defining a map $\gamma_f : S_0 \to S_0$.

DEFINITION 2.6.1. The *index of the point p with respect to the curve f* is defined as the degree of the map γ, i.e.,

$$\mathrm{ind}(p, f) := \deg(\gamma_f)$$

Clearly, $\mathrm{ind}(p, f)$ does not change when p varies inside a connected component of $\mathbb{R}^2 \backslash C$): indeed, the function ind is continuous in $p \notin C$ and takes integer values, so it has to be a constant when p varies in a connected component of $\mathbb{R}^2 \backslash C$).

If the point p is "far from" $f(S^1)$ (i.e., in the connected component of $\mathbb{R}^2 \backslash f(\mathbb{S}^1)$ with noncompact closure), then $\deg(p, f) = 0$; indeed, if p is sufficiently far from C (which is compact), then C is contained in an acute angle with vertex at p, so that the vector $f(\varphi)$ remains within that angle as φ varies from 0 to 2π and γ must have degree 0.

A concrete example of a curve in $\mathbb{R}^2$ is shown on Figure ??, (a); on it, the integers indicate the values of the index in each connected component of its complement.

is this 2.6.1?; v

2.6.2. Computing the index for immersed curves.

When the curve is nice enough, there is a convenient method for computing the index of any point with respect to the curve. To formalize what we mean by "nice" we introduce the following definition.

DEFINITION 2.6.2. A curve $f : \mathbb{S}^1 \to \mathbb{R}^2$ is said to be an *immersion* if f is differentiable, has a nonzero tangent vector, and has a finite number of self-intersections, all of them transversal, i.e. with all tangent vectors making non-zero angles with each other.

In order to compute the index of p with respect to an immersed curve f, let us join p by a (nonclosed) smooth curve α transversal to f to a far away point a and move from a to p along that curve. At the start, we put $i(a) = 0$, and, moving along α, we add one to i when we cross $f(\mathbb{S}^1)$ in the positive direction (i.e., so that the tangent vector to f looks to the right of α) and subtract one when we cross it in the negative direction. When we reach the connected component of the complement to the curve containing p, we will obtain a certain integer $i(p)$.

EXERCISE 2.6.1. Prove that the integer $i(p)$ obtained in this way is actually the index of p w.r.t. f (and so $i(p)$ does not depend on the choice of the curve α).

Turn this into a pr
pro

EXERCISE 2.6.2. Compute the indices of the connected components of the complements to the curve shown on Figure ??(b) by using the algorithm described above.

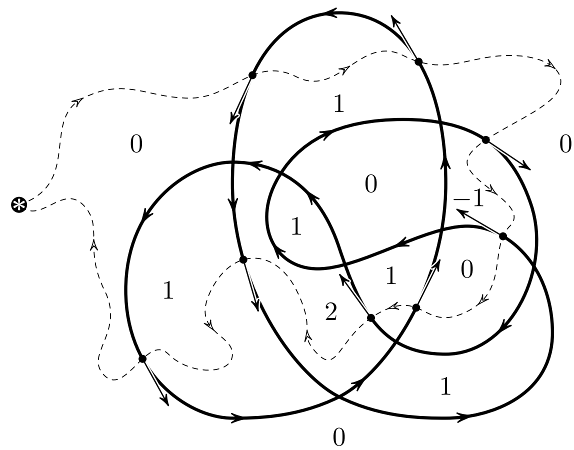

FIGURE 2.6.1. Index of points w.r.t. a curve

2.7. The fundamental theorem of algebra

2.7.1. Statement and commentary. In our times the term "fundamental theorem of algebra" reflects historical preoccupation of mathematicians with solving algebraic equations, i.e. finding roots of polynomials. Its equivalent statement is that the field of complex numbers is *algebraically complete* i.e. that no need to extend it in order to perform algebraic operations. This in particular explain difficulties with constructing "hyper-complex" numbers; in order to do that in a meaningful way, one needs to relax some of the axioms of the field (e.g. commutativity for the four-dimensional *quaternions*).[3] Thus, in a sense, the theorem is fundamental but not so much for algebra where the field of complex numbers is only one of many objects of study, and not the most natural one at that, but for analysis, analytic number theory and classical algebraic geometry.

THEOREM 2.7.1. *Any polynomial*

$$p(z) = a_n z^n + a_{n-1}z^{n-1} + \cdots + a_1 z + a_0, \quad a_n \neq 0, \quad n > 0,$$

with complex coefficients has a least one complex root.[4]

REMARK 2.7.2. This theorem has many different proofs, but no "purely algebraic" ones. In all existing (correct!) proofs, the crucial point is topological. In the proof given below, it ultimately comes down to the fact that a degree n self map of the circle is not homotopic to the identity provided $n \geq 2$.

[3]This by no means implies that one cannot include complex number to a larger field. General algebraic constructions such as fields of rational functions provide for that.

[4]The fact that the polynomial $x^2 + 1$ has no *real* roots is the most basic motivation for introducing complex numbers.

2.7.2. Proof of the theorem. By dividing all coefficients by a_n which does not change the roots we may assume that $a_n = 1$. Furthermore, if $a_0 = 0$ than $p(0) = 0$. Thus we can also assume that $a_0 \neq 0$.

Consider the curve $f_n : S^1 \to \mathbb{R}^2$ given by the formula $e^{i\varphi} \mapsto R_0^n e^{in\varphi}$, where R_0 is a (large) positive number that will be fixed later. Further, consider the family of curves $f_{p,R} : S^1 \to \mathbb{R}^2$ given by the formula $e^{i\varphi} \mapsto p(Re^{i\varphi})$, where $R \leq R_0$. We can assume that the origin O does not belong to $f_{p,R_0}(\mathbb{S}^1)$ (otherwise the theorem is proved).

LEMMA 2.7.3. *If R_0 is sufficiently large, then*

$$\operatorname{ind}(O, f_{p,R_0}) = \operatorname{ind}(O, f_n) = n.$$

Before proving the lemma, let us show that it implies the theorem.

By the lemma, $\operatorname{ind}(O, f_{p,R_0}) = n$. Let us continuously decrease R from R_0 to 0. If for some value of R the curve $f_{p,R}(\mathbb{S}^1)$ passes through the origin, the theorem is proved. So we can assume that $\operatorname{ind}(O, f_{p,R})$ changes continuously as $R \to 0$; but since the index is an integer, it remains constant and equal to n. However, if R is small enough, the curve $f_{p,R}(\mathbb{S}^1)$ lies in a small neighborhood of a_0; but for such an R we have $\operatorname{ind}(O, f_{p,R}) = 0$. This is a contradiction, because $n \geq 1$.

It remains to prove the lemma. The equality $\operatorname{ind}(O, f_n) = n$ is obvious. To prove the other equality, it suffices to show that for any φ the difference Δ between the vectors $V_p(\varphi)$ and $V_n(\varphi)$ that join the origin O with the points $f_p(R_0e^{i\varphi})$ and $f_n(R_0e^{i\varphi})$, respectively, is small in absolute value (as compared to $R_0^n = |V_p(\varphi)|$) if R_0 is large enough. Indeed, by the definition of degree, if the mobile vector is replaced by another mobile vector whose direction always differs from the direction of the first one by less than $\pi/2$, the degree will be the same for the two vectors.

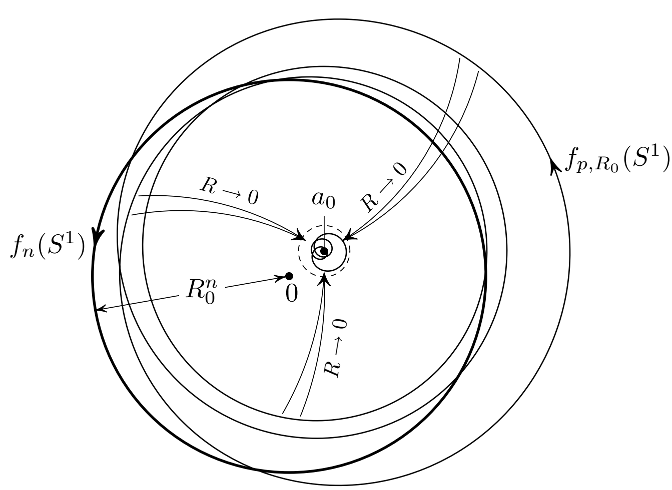

FIGURE 2.7.1. Proof of the fundamental theorem of algebra

Clearly, $|\Delta| = |a_{n-1}z^{n-1} + \cdots + a_1 z + a_0|$. Let us estimate this number, putting $z = R_0 e^{\varphi}$ and $A = \max\{a_{n-1}, a_{n-2}, \ldots, a_0\}$ (here without loss of generality we assume that $R_0 > 1$). We then have

$$|\Delta| = |a_{n-1}z^{n-1} + \cdots + a_1 z + a_0| \le |A(R_0^{n-1} + R_0^{n-2} \cdots + 1)| \le A \cdot n \cdot R_0^{n-1}.$$

Now if we put $R_0 := K \cdot A$, where K is a large positive number, we will obtain $|\Delta| \le nA(KA)^{n-1} = nK^{n-1}A^n$. Let us compare this quantity to R_0^n; the latter equals $R_0^n = K^n A^n$, so for K large enough the ratio $|\Delta|/R_0^n$ is as small as we wish. This proves the lemma and concludes the proof of the theorem. □

2.8. The fundamental group; definition and elementary properties

ns for Mobius and (the latter need to fined in Chapter 1

The fundamental group is one of the most important invariants of homotopy theory. It also has numerous applications outside of topology, especially in complex analysis, algebra, theoretical mechanics, and mathematical physics. In our course, it will be the first example of a "functor", assigning a group to each path-connected topological space and a group homomorphism to each continuous map of such spaces, thus reducing topological problems about spaces to problems about groups, which can often be effectively solved. In a more down-to-earth language this will be the first sufficiently universal non-trivial invariant of homotopy equivalence, defined for all path connected spaces and calculable in many natural situations.

2.8.1. Main definitions. Let M be a topological space with a marked point $p \in M$.

DEFINITION 2.8.1. A curve $c\colon [0,1] \to M$ such that $c(0) = c(1) = p$ will be called a *loop* with *basepoint* p. Two loops c_0, c_1 with basepoint p are called *homotopic rel* p if there is a homotopy $F : [0,1] \times [0,1] \to M$ joining c_0 to c_1 such that $F(t, x) = p$ for all $t \in [0,1]$.

If c_1 and c_2 are two loops with basepoint p, then the loop $c_1 \cdot c_2$ given by

$$c_1 \cdot c_2(t) := \begin{cases} c_1(2t) & \text{if } t \le \frac{1}{2}, \\ c_2(2t-1) & \text{if } t \ge \frac{1}{2}. \end{cases}$$

is called the *product* of c_1 and c_2.

PROPOSITION 2.8.2. *Classes of loops homotopic rel p form a group with respect to the product operation induced by $\cdot$.*

PROOF. First notice that the operation is indeed well defined on the homotopy classes. For, if the paths c_i are homotopic to $\tilde{c}_i$, $i = 1, 2$ via the maps $h_1 : [0,1] \times [0,1] \to M$, then the map h, defined by

$$h(t,s) := \begin{cases} h_1(2t, s) & \text{if } \ t \le \frac{1}{2}, \\ h_2(2t-1, s) & \text{if } t \ge \frac{1}{2} \end{cases}$$

is a homotopy rel p joining c_1 to c_2.

Obviously, the role of the unit is played by the homotopy class of the constant map $c_0(t) = p$. Then the inverse to c will be the homotopy class of the map $c'(t) := c(1-t)$. What remains is to check the associative law: $(c_1 \cdot c_2) \cdot c_3$ is homotopic rel p to $c_1 \cdot (c_2) \cdot c_3)$ and to show that $c \cdot c'$ is homotopic to c_0. In both cases the homotpy is done by a reparametrization in the preimage, i.e., on the square $[0,1] \times [0,1]$.

For associativity, consider the following continuous map ("reparametrization") of the square into itself

$$R(t,s) = \begin{cases} (t(1+s), s) & \text{if } 0 \le t \le \frac{1}{4}, \\ (t + \frac{s}{4}, s) & \text{if } \frac{1}{4} \le t \le \frac{1}{2}, \\ (1 - \frac{1}{1+s} + \frac{t}{1+s}, s) & \text{if } \frac{1}{2} \le t \le 1. \end{cases}$$

Then the map $c_1 \cdot (c_2 \cdot c_3) \circ R : [0,1] \times [0,1] \to M$ provides a homotopy rel endpoints joining the loops $c_1 \cdot (c_2 \cdot c_3)$ and $(c_1 \cdot c_2) \cdot c_3$.

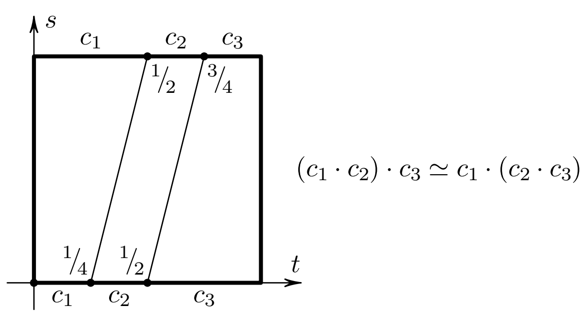

FIGURE 2.8.1. Associativity of multiplication

Similarly, a homotopy joining $c \cdot c'$ to c_0 is given by $c \cdot c' \circ I$, where the reparametrization $I : [0,1] \times [0,1] \to [0,1] \times [0,1]$ is defined as

$$I(t,s) = \begin{cases} (t,s) & \text{if } 0 \le t \le \frac{1-s}{2}, \text{ or } \frac{1+s}{2} \le t \le 1, \\ (\frac{1-s}{2}, s) & \text{if } \frac{1-s}{2} \le t \le \frac{1+s}{2}, \end{cases}$$

Notice that while the reparametrization I is discontinuous along the wedge $t = (1 \pm s)/2$, the map $(c \cdot c') \circ I$ is continuous by the definition of c'. $\square$

DEFINITION 2.8.3. The group described in Proposition 2.8.2 is called the *fundamental group* of M at p and is denoted by $\pi_1(M,p)$.

It is natural to ask to what extent $\pi_1(M,p)$ depends on the choice of the point $p \in M$. The answer is given by the following proposition.

PROPOSITION 2.8.4. *If p and q belong to the same path connected component of M, then the groups $\pi_1(M,p)$ and $\pi_1(M,q)$ are isomorphic.*

PROOF. Let $\rho : [0,1] \to M$ be a path connecting points p and q. It is natural to denote the path $\rho \circ S$ where $S(t) = 1 - t$ by ρ^{-1}. It is also natural to extend the " $\cdot$ " operation to paths with different endpoints if they match properly. With these conventions established, let us associate to a path $c\colon [0,1] \to M$ with $c(0) = c(1) = p$ the path $c' := \rho^{-1} \cdot c \cdot \rho$ with $c'(0) = c'(1) = q$. In order to finish the proof, we must show that this correspondence takes paths homotopic rel p to paths homotopic rel q, respects the group operation and is bijective up to homotopy. These staments are proved using appropriate rather natural reparametrizations, as in the proof of Proposition 2.8.2. □

REMARK 2.8.5. By mapping the interval $[0,1]$ to the circle with a marked point e first and noticing that, if the endpoints are mapped to the e, than the homotopy can also be interpreted as a map of the closed cylinder $\mathbb{S}^1 \times [0,1]$ to the space with a based point which maps $e \times [0,1]$ to the base point p, we can interpret the construction of the fundamental group as the group of homotopy classes of maps $(\mathbb{S}^1, e)$ into $M, p)$. Sometimes this language is more convenient and we will use both versions interchangeably.

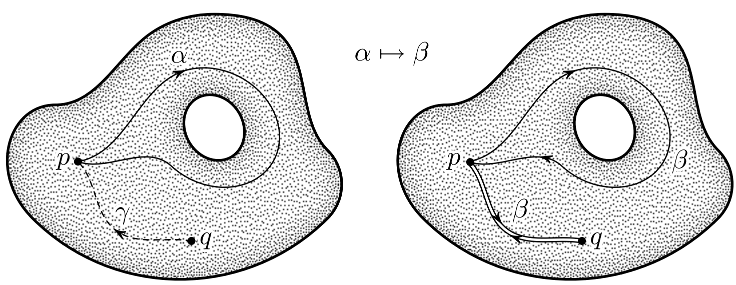

FIGURE 2.8.2. Change of basepoint isomorphism

REMARK 2.8.6. It follows from the construction that different choices of the connecting path ρ will produce isomorphisms between $\pi_1(M,p)$ and $\pi_1(M,q)$ which differ by an inner automorphism of either group.

If the space M is path connected then the fundamental groups at all of its points are isomorphic and one simply talks about the *fundamental group of* M and often omits the basepoint from its notation: $\pi_1(M)$.

DEFINITION 2.8.7. A path connected space with trivial fundamental group is said to be *simply connected* (or sometimes 1-*connected*).

REMARK 2.8.8. Since the fundamental group is defined modulo homotopy, it is the same for homotopically equivalent spaces, i.e., it is a homotopy invariant.

The *free* homotopy classes of curves (i.e., with no fixed base point) correspond exactly to the conjugacy classes of curves modulo changing base point, so there is a natural bijection between the classes of freely homotopic closed curves and conjugacy classes in the fundamental group.

That this object has no natural group structure may sound rather unfortunate to many a beginner topologist since the main tool of algebraic topology, namely, translating difficult geometric problems into tractable algebraic ones, have to be applied here with fair amount of care and caution here.

2.8.2. Functoriality. Now suppose that X and Y are path connected, $f : X \to Y$ is a continuous maps with and $f(p) = q$. Let $[c]$ be an element of $\pi_1(X, p)$, i.e., the homotopy class rel endpoints of some loop $c : [0, 1] \to X$. Denote by $f_\#(c)$ the loop in (Y, q) defined by $f_\#(t) := f(c(t))$ for all $t \in [0, 1]$. $\|f_*^n(v)\| \geq Ce^{-\mu|n|}\|v\|, \quad \text{for all} \quad n > 0 \quad \text{and} \quad v \in E^c(p)$. The following simple but fundamental fact is proven by a straightforward checking that homotopic rel based points loop define homotopic images.

PROPOSITION 2.8.9. *The assignment $c \mapsto f_\#(c)$ is well defined on classes of loops and determines a homomorphism (still denoted by $f_\#$) of fundamental groups:*

$$f_\# : \pi_1(X, p) \to \pi_1(Y, q)$$

(refered to as the homomorphism induced by f), which possesses the following properties (called functorial):

- $(f \circ g)_\# = f_\# \circ g_\#$ *(covariance);*
- *$(id_X)_\# = id_{\pi_1(X,p)}$ (identity maps induce identity homomorphisms).*

The fact that the construction of an invariant (here the fundamental group) is functorial is very convenient for applications. For example, let us give another proof of the Brouwer fixed point theorem for the disk by using the isomorphism $\pi_1(\mathbb{S}^1) = \mathbb{Z}$ (see Proposition 2.8.12 below) and $\pi_1(\mathbb{D}^2) = 0$ (since $\mathbb{D}^2$ is contractible) and the functoriality of $\pi_1(\cdot)$.

We will prove (by contradiction) that there is no *retraction* of $\mathbb{D}^2$ on its boundary $\mathbb{S}^1 = \partial\mathbb{D}^2$ i.e. a map $\mathbb{D}^2 \to \mathbb{S}^1$ which is identity on $\mathbb{S}^1$, Let $r : \mathbb{D}^2 \to \mathbb{S}^1$ be such a retraction, let $i : \mathbb{S}^1 \to \mathbb{D}^2$ be the inclusion; choose a basepoint $x_0 \in \mathbb{S}^1 \subset \mathbb{D}^2$. Note that for this choice of basepoint we have $i(x_0 = r(x_0) = x_0)$. Consider the sequence of induced maps:

$$\pi_1(\mathbb{S}^1, x_0) \xrightarrow{i_*} \pi_1(\mathbb{D}^2, x_0) \xrightarrow{r_*} \pi_1(\mathbb{S}^1, x_0).$$

In view of the isomorphisms noted above, this sequence is actually

$$\mathbb{Z} \xrightarrow{i_*} 0 \xrightarrow{r_*} \mathbb{Z}.$$

But such a sequence is impossible, because by functoriality we have

$$r_* \circ i_* = (r \circ i)_* = \mathrm{Id}_* = \mathrm{Id}_{\mathbb{Z}}\,. \quad \square$$

In addition to functoriality the fundamental group behaves nicely with respect to the product.

PROPOSITION 2.8.10. *If X and Y are path connected spaces, then*

$$\pi_1(X \times Y) = \pi_1(X) \times \pi_1(Y).$$

PROOF. Let us construct an isomorphism of $\pi_1(X) \times \pi_1(Y)$ onto $\pi_1(X \times Y)$. Let x_0, y_0 be the basepoints in X and Y, respectively. For the basepoint in $X \times Y$, let us take the point (x_0, y_0). Now to the pair of loops α and β in X and Y let us assign the loop $\alpha \times \beta$ given by $\alpha \times \beta(t) := (\alpha(t), \beta(t))$. The verification of the fact that this assignment determines a well-defined isomorphism of the appropriate fundamental groups is quite straightforward. For example, to prove surjectivity, for a given loop γ in $X \times Y$ with basepoint (x_0, y_0), we consider the two loops $\alpha(t) := (\mathrm{pr}_X \circ \gamma)(t)$ and $\beta(t) := (\mathrm{pr}_Y \circ \gamma)(t)$, where pr_X and pr_Y are the projections on the two factors of $X \times Y$. □

COROLLARY 2.8.11. *If C is contractible, then $\pi_1(X \times C) = \pi_1(X)$*

EXERCISE 2.8.1. Prove that for any path connected topological space X we have $\pi_1(\text{Cone(X)}) = 0$.

2.8.3. Examples and applications. The first non-trivial example is an easy corollary of degree theory.

PROPOSITION 2.8.12. *The fundamental group of the circle $\mathbb{R}/\mathbb{Z}$ is $\mathbb{Z}$ and in additive notation for $\mathbb{S}^1 = \mathbb{R}/\mathbb{Z}$ with 0 being the base point the element $n \in \mathbb{Z}$ is represented by the map E_m.*

PROOF. This is essentially a re-statement of Theorem 2.4.4. Since this is a very fundamental fact of homotopy theory we give a detailed argument.

Let $\gamma : (\mathbb{S}^1, 0) \to (\mathbb{S}^1, 0)$ be a loop. LIft it in a unique fashion to a map $\Gamma : (\mathbb{R}, 0) \to \mathbb{R}, 0)$. A homotopy rel 0 between any two maps $\gamma, \gamma' : (\mathbb{S}^1, 0) \to (\mathbb{S}^1, 0)$ lifts uniquely between a homotopy between lifts. Hence $\deg\gamma$ is a homotopy invariant of γ. On the other hand the "straight-line homotopy" between Γ and the linear map $x \deg \gamma x$ projects to a homotopy rel 0 between γ and $E_{\deg\gamma}$. □

Proposition 2.8.12 and Proposition 2.8.10 immediately imply

COROLLARY 2.8.13. $\pi_1(\mathbb{T}^n) = \mathbb{Z}^n$.

Notice that $\mathbb{T}^n = \mathbb{R}^n/\mathbb{Z}^n$ and $\pi_1(\mathbb{T}^n)$ is isomorphic to the subgroup $\mathbb{Z}^n$ by which $\mathbb{R}^n$ is factorized. This is not accidental but the first instance of *universal covering* phenomenon, see Section 6.2.2.

On the other hand here is an example of a space, which later will be shown to be non-contractible, with trivial fundamental group.

PROPOSITION 2.8.14. *For any $n \geq 2$, $\pi_1(\mathbb{S}^n) = 0$.*

PROOF. The main idea of the proof is to make use of the fact that $\mathbb{S}^n$ is the one-point compactification of the contractible space $\mathbb{R}^n$ and that for $n \geq 2$ any loop is homotopic to one which avoids this single point. For such a loop the contraction (deformation) of $\mathbb{S}^n$ with one point removed to the base point of the loop also produces a homotopy of the loop to the trivial one. However exotic a loop whose image covers the whole sphere may look such loops exist (see Peano curves, **??**). Still any loop is homotopic to a loop which consist of a finite number of arcs of great circles and hence does not cover the whole sphere. The method we use here is interesting since we will make use of a *geometric* structure (spherical geometry on this occasion) to prove a purely topological statement, so we will describe it in detail.

For any two points p and q on the standard unit n-sphere in $\mathbb{R}^{n+1}$, which are not diametrically opposite, there is a unique shortest curve connecting these points, namely the shorter of the two arcs of the *great circle* which can be described as the intersection of the two-dimensional plane passing through p, q and the origin. Such curves give the next simplest example after straight lines in the Euclidean space of *geodesics* which play a central role in Riemannian geometry, the core part of differential geometry. We will mention that subject somewhat more extensively in **??** and will describe the basics of a systematic theory in **??**. An important thing to remember is that any geodesic is provided with the natural *length parameter* and that they depend continuously on the endpoints as long those are not too far away (e.g. are not diametrically opposite in the case of the standard round sphere).

Now come back to our general continuous loop γ in $\mathbb{S}^n$. By compactness one can find finitely many points $0 = t_0 < t_1 < \cdots < t_{m-1}, t_m = 1$ such that for $k = 0, 1 \ldots, m-1$ the set $\Gamma_k := \gamma[t_k, t_{k+1}]$ lies is a sufficiently small ball. In fact for our purpose it would be sufficient if this set lies within an open half-sphere. Now for any open half-sphere $H \subset \mathbb{S}^n$ and any $p, q \in H$ there is a canonical homotopy of H into the arc of the great circle C in H connecting p and q keeping these two points fixed. Namely, first for any $x \in H$ consider the unique arc A_x of the great circle perpendicular to the great circle C and connecting x with C and lying in H. Our homotopy moves x along A_x according to the length parameter normalized to 1/2. The result is a homotopy of H to $C \cap H$ keeping every point on $C \cap H$ fixed. After that one contracts $C \cap H$ to the arc between p and q by keeping all points on that arc fixed and uniformly contracting the length parameter normalized to 1/2 on the remaining two arcs. This procedure restricted to $\gamma[t_k, t_{k+1}]$ on each interval $[t_k, t_{k+1}]$, $k = 0, \ldots, m-1$ produces a homotopy of γ to a path whose image is a finite union of arcs of great circles and hence does not cover the whole sphere. $\square$

it would be
illus

Now we can make an advance toward a solution of a natural problem which concerned us since we first introduced manifolds: invariance of dimension. We proved that one-dimensional manifolds and higher dimensional ones are not homeomorphic by an elementary observation that removing a single point make the former disconnected locally while the latter remains connected. Now we can make a step forward from one to two. This will be the first instance when we prove absence of homeomorphism by appealing to homotopy equivalence.

PROPOSITION 2.8.15. *Any two-dimensional manifold and any n-dimensional manifold for $n \geq 3$ are not homeomorphic.*

PROOF. First let us show that $\mathbb{R}^2$ and $\mathbb{R}^n$ for $n \geq 3$ are not homeomorphic. By removing one point we obtain in the first case the space homotopically equivalent to the circle which hence has fundamental group $\mathbb{Z}$ by Proposition 2.8.12 and in the second the space homotopically equivalent to $\mathbb{S}^{n-1}$ which is simply connected by Proposition 2.8.14.

Now assume that $h : M^n \to M^2$ is a homeomorphism from an n-dimensional manifold to a two-dimensional manifold. Let $h(p) = q$. Point p has a base of neighborhoods homeomorphic to $\mathbb{R}^n$. Hence any loop in such a neighborhood which does not touch p can be contracted to a point within the neighborhood without the homotopy touching p. On the other hand, q has a base of neighborhoods homeomorphic to $\mathbb{R}^2$ which do not possess this property. Let $N \ni q$ be such a neighborhood and let $N' \ni p$ be a neighborhood of p homeomorphic to $\mathbb{R}^n$ such that $h(N') \supset N$. Let $\gamma : [0,1] \to N' \setminus \{p\}$ be a loop which is hence contractible in $N' \setminus \{p\}$. Then $h \circ \gamma : [0,1] \to h(N') \setminus \{q\}$ is a loop which is contractible in $h(N') \setminus \{q\}$ and hence in $N \setminus \{q\}$, a contradiction. $\square$

REMARKS 2.8.16.
(1) In order to distinguish between the manifolds of dimension higher than two the arguments based on the fundamental group are not sufficient. One needs either higher homotopy group s introduced below in Section 2.10 or degree theory for maps of spheres of higher dimension **??**.
(2) Our argument above by no means shows that manifolds of different dimension are not homotopically equivalent; obviously all $\mathbb{R}^n$s are since they are all contractible. More interestingly even, the circle and Móbius strip are homotopically equivalent as we already know. However a proper even more general version of degree theory (which is a basic part of homology theory for manifolds) will allow as to show that dimension is an invariant of homotopy equivalence for *compact* manifolds.

2.8.4. The Seifert–van Kampen theorem. In this subsection we state a classical theorem which relates the fundamental group of the union of two spaces with the fundamental groups of the summands and of their intersection. The result turns out to give an efficient method for computing the fundamental group of a "complicated" space by putting it together from "simpler" pieces.

In order to state the theorem, we need a purely algebraic notion from group theory.

DEFINITION 2.8.17. Let G_i, $i = 1, 2$, be groups, and let $\varphi_i : K \to G_i$, $i = 1, 2$ be monomorphisms. Then the *free product with amalgamation* of G_1 and G_2 with respect to φ_1 and φ_2, denoted by $G_1 *_K G_2$ is the quotient group of the free product $G_1 * G_2$ by the normal subgroup generated by all elements of the form $\varphi_1(k)(\varphi_2(k))^{-1}$, $k \in K$.

THEOREM 2.8.18 (Van Kampen's Theorem). *Let the path connected space X be the union of two path connected spaces A and B with path connected intersection containing the basepoint $x_0 \in X$. Let the inclusion homomorphisms*

$$\varphi_A : \pi_1(A \cap B) \to \pi_1(A), \quad \varphi_B : \pi_1(A \cap B) \to \pi_1(B)$$

be injective. Then $\pi_1(X, x_0)$ is the amalgamated product

$$\pi_1(X, x_0) \cong \pi_1(A, x_0) *_{\pi_1(A \cap B, x_0)} \pi_1(B, x_0).$$

For a proof see G, Bredon, Geometry and Topology, Theorem 9.4.

2.9. The first glance at covering spaces

A covering space is a mapping of spaces (usually manifolds) which, locally, is a homeomorphism, but globally may be quite complicated. The simplest nontrivial example is the exponential map $\mathbb{R} \to \mathbb{S}^1$ discussed in Section 2.4.1.

2.9.1. Definition and examples.

DEFINITION 2.9.1. If M, M' are topological manifolds and $\pi \colon M' \to M$ is a continuous map such that $\operatorname{card} \pi^{-1}(y)$ is independent of $y \in M$ and every $x \in \pi^{-1}(y)$ has a neighborhood on which π is a homeomorphism to a neighborhood of $y \in M$ then π is called a *covering map* and M' (or (M', π)) is called a *covering (space)* or *cover* of M. If $n = \operatorname{card} \pi^{-1}(y)$ is finite, then (M', π) is said to be an *n-fold covering*.

If $f \colon N \to M$ is continuous and $F \colon N \to M'$ is such that $f = \pi \circ F$, then F is said to be a *lift* of f. If $f \colon M \to M$ is continuous and $F \colon M' \to M'$ is continuous such that $f \circ \pi = \pi \circ F$ then F is said to be a *lift* of f as well.

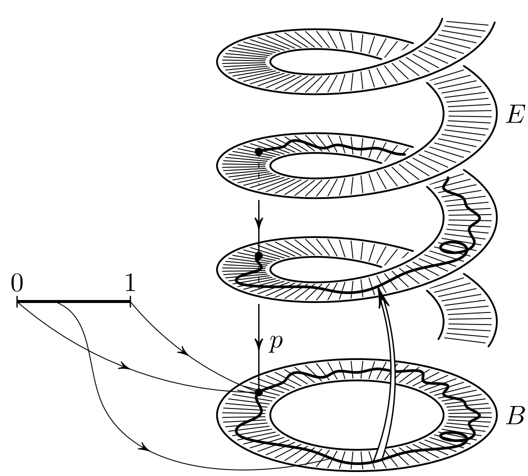

FIGURE 2.9.1. Lift of a closed curve

DEFINITION 2.9.2. A simply connected covering is called the *universal cover*. A homeomorphism of a covering M' of M is called a *deck transformation* if it is a lift of the identity on M.

EXAMPLE 2.9.3. $(\mathbb{R}, \exp(2\pi i(\cdot)))$ is a covering of the unit circle. Geometrically one can view this as the helix $(e^{2\pi i x}, x)$ covering the unit circle under projection. The map defined by taking the fractional part likewise defines a covering of the circle $\mathbb{R}/\mathbb{Z}$ by $\mathbb{R}$.

PROPOSITION 2.9.4. *If* $\pi\colon M' \to M$ *and* $\rho\colon N' \to N$ *are covering maps, then* $\pi \times \rho\colon M' \times N' \to M \times N$ *is a covering map.*

EXAMPLE 2.9.5. The torus $\mathbb{T}^2 = \mathbb{S}^1 \times \mathbb{S}^1$ is covered by the cylinder $\mathbb{S}^1 \times \mathbb{R}$ which is in turn covered by $\mathbb{R}^2$. Notice that the fundamental group $\mathbb{Z}$ of the cylinder is a subgroup of that of the torus ($\mathbb{Z}^2$) and $\mathbb{R}^2$ is a simply connected cover of both.

EXAMPLE 2.9.6. The maps $E_m,\ |m| \geq 2$ of the circle define coverings of the circle by itself.

EXAMPLE 2.9.7. The natural projection $\mathbb{S}^n \to \mathbb{R}P(n)$ which send points x and $-x$ into their equivalence class is a two-fold covering. On the other, hand, the identification map $\mathbb{S}^{2n-1} \to \mathbb{C}P(n)$ is not a covering since the pre-image on any point is a continuous curve.

EXERCISE 2.9.1. Describe two-fold coverings of
(1) the (open) Möbius strip by the open cylinder $\mathbb{S}^1 \times \mathbb{R}$;
(2) the Klein bottle by the torus $\mathbb{T}^2$.

2.9.2. Role of the fundamental group. One of the remarkable aspects of any covering space $p : X \to B$ is that it is, in a sense, entirely governed by the fundamental groups of the spaces B and X, or more precisely, by the induced homomorphism $p_\# : \pi_1(X) \to \pi_1(B)$ of their fundamental groups. We shall observe this in the two examples given below, postponing the exposition of the general theory to Chapter 6.

EXAMPLE 2.9.8. Let B be the plane annulus given by the inequalities $1 \leq r \leq 2$ in the polar coordinates (r, φ) on the plane $\mathbb{R}^2$, and let X be another copy of this annulus. Consider the map $p : X \to B$ given by $(r, \varphi) \mapsto (r, 3\varphi)$. It is obviously a covering space. Geometrically, it can be viewed as in the figure, i.e., as the vertical projection of the strip $aba'b'$ (with the segments ab and $a'b'$ identified) onto the horizontal annulus.

FIGURE ?? A triple covering of the annulus

The fundamental group of B (and of X) is isomorphic to $\mathbb{Z}$, and the induced homomorphism $p_\# : \pi_1(X) \to \pi_1(B)$ is the monomorphism of $\mathbb{Z}$ into $\mathbb{Z}$ with image $3\mathbb{Z} \subset \mathbb{Z}$. The deck transformations constitute a group isomorphic to $\mathbb{Z}_3 \cong \mathbb{Z}/3\mathbb{Z}$.

This is a fairly general situation. The homomorphism $p_\#$ is always injective (for any covering space p) and, provided $\mathrm{Im}(p_\#)$ is a normal subgroup of $\pi_1(B)$, [5] the deck transformations form a group isomorphic to the quotient $\pi_1(B)\big/\mathrm{Im}(p_\#)$.

More remarkable is that the covering map p is entirely determined (up equivalence, defined in a natural way) by the choice of a subgroup of $\pi_1(B)$, in our case, of the infinite cyclic subgroup of $\pi_1(B)$ generated by the element $3e$, where e is the generator of $\pi_1(B) \cong \mathbb{Z}$. There is in fact a geometric procedure for constructing the covering space X, which in our case will yield the annulus.

Another way of defining the geometric structure of a covering space in algebraic terms is via the action of a discrete group in some space X. Then the covering is obtained as the quotient map of X onto the orbit space of the group action. In our case the space X is the annulus, the discrete group is $\mathbb{Z}_3$ and it acts on X by rotations by the angles 0, $2\pi/3$, $4\pi/3$, the orbit space is B (another annulus), and the quotient map is p.

EXAMPLE 2.9.9. Let B be the torus $\mathbb{S}^1 \times \mathbb{S}^1$ with coordinates (φ, ψ) and X be the cylinder $r = 1$ in 3-space endowed with the cylindrical coordinates (r, θ, h). Consider the map $p : X \to B$ given by

$$(1, \varphi, h) \mapsto (2\varphi,\ h \mod 2\pi).$$

It is obviously a covering space map. Geometrically, it can be described as wrapping the cylinder an infinite number of times along the parallels of the torus and simultaneously covering it twice along the meridians.

The fundamental group of B is isomorphic to $\mathbb{Z}$, that of X is $\mathbb{Z} \oplus \mathbb{Z}$ and the induced homomorphism $p_\# : \pi_1(X) \to \pi_1(B)$ is the monomorphism of $\mathbb{Z}$ into $\mathbb{Z} \oplus \mathbb{Z}$ with image $2\mathbb{Z} \oplus \mathbb{Z} \subset \mathbb{Z} \oplus \mathbb{Z}$. The deck transformations constitute a group isomorphic to $(\mathbb{Z}/2\mathbb{Z}) \oplus \mathbb{Z} \subset \mathbb{Z} \oplus \mathbb{Z}$.

Here also the covering p can be obtained by an appropriate choice of a discrete group acting on the cylinder X; then p will be the quotient map of X onto the orbit space of this action.

For an arbitrary "sufficiently nice" space B, say a manifold, there is natural bijection between conjugacy classes of subgroups of $\pi_1(M)$ and classes of covering spaces modulo homeomorphisms commuting with deck transformations. This bijection will be described in detail in Chapter 6, where it will be used, in particular, to prove the uniqueness of the universal cover.

[5] This is an important condition which prevents pathologies which may appear for other coverings

2.10. Definition of higher homotopy groups

The fundamental group has natural generalizations (with $\mathbb{S}^1$ replaced by $\mathbb{S}^n$, $n \geq 2$) to higher dimensions, called (higher)*homotopy groups* (and denoted by $\pi_n(\cdot)$). The higher homotopy groups are just as easy (in a sense easier) to define than the fundamental group, and, unlike the latter, they are commutative.

Let X be a topological space with a marked point $p \in X$. On the sphere $\mathbb{S}^n$, fix a marked point $q \in \mathbb{S}^n$, and consider a continuous map

$$f : \mathbb{S}^n \to X \quad \text{such that} \quad f(q) = p.$$

Such a map is called a *spheroid*. Two spheroids are considered *equivalent* if they are homotopic rel basepoints, i.e., if there exists a homotopy $h_t : \mathbb{S}^n \to X$, $t \in [0, 1]$, joining the two spheroids and satisfying $h_t(q) = p$ for all $t \in [0, 1]$. By an abuse of language, we will also refer to the corresponding equivalence classes as spheroids. It is sometimes more convenient to regard spheroids as homotopy classes of maps

$$f : (\mathbb{D}^n, \partial\mathbb{D}^n) \to (X, p),$$

where the homotopy h_t must take $\partial\mathbb{D}^n$, the $n-1$-dimensional sphere $\mathbb{S}^{n-1}$, to p for all $t \in [0, 1]$.

Let us denote by $\pi_n(X, p)$ the set of all (equivalence classes of) spheroids and introduce a binary operation in that set as follows. Suppose $f, g : (\mathbb{S}^n, q) \to (X, p)$ are two spheroids; then their product is the spheroid $fg : (\mathbb{S}^n, q) \to (X, p)$ obtained by pulling the equator of $\mathbb{S}^n$ containing p to a point and then defining fg by using f on one of the two spheres in the obtained wedge and g on the other (see the figure).

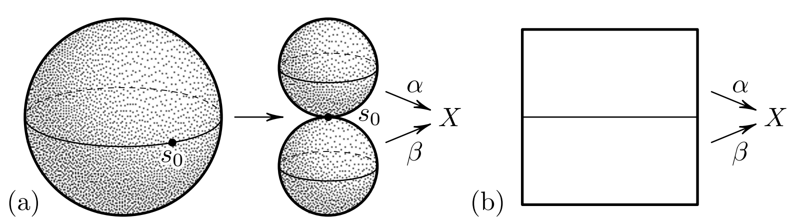

FIGURE 2.10.1. The product of two spheroids

Note that for $n = 1$ this definition coincides with the product of loops for the fundamental group $\pi_1(X, p)$. We will also sometimes consider the set $\pi_0(X, p)$, which by definition consists of the path connected components of X and has no natural product operationdefined on it.

PROPOSITION 2.10.1. *For $n \geq 2$ and all path connected spaces X, the set $\pi_n(X, p)$ under the above definition of product becomes an Abelian group, known as the n-th* homopoty group of X with basepoint p.

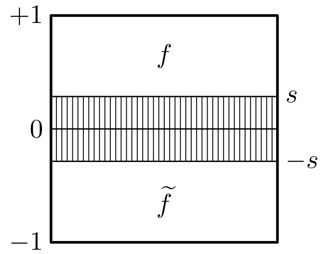

FIGURE 2.10.2. Inverse element in $\pi_n(X, p)$

PROOF. The verification of the fact that $\pi_n(X, p))$ is a group is straightforward; we will only show how inverse elements are constructed. This construction is shown on the figure.

On the figure $f : (\mathbb{D}^n, \partial\mathbb{D}^n) \to (X, p)$ is a spheroid. Denote by $\widetilde{f} : I^{n-1} \times [-1, 1] \to X$ the spheroid given by $\widetilde{f}(x, s) := f(x, -s)$. Then the map (spheroid) $f\widetilde{f}$ satisfies $f\widetilde{f}(x, s) = \widetilde{f}f(x, -s)$ (look at the figure again). Therefore we can consider the family of maps

$$h_t(x, s) = \begin{cases} f\widetilde{f}(x, s), & \text{for } |s| \geq t, \\ \widetilde{f}f(x, -s), & \text{for } |s| \leq t. \end{cases}$$

For this family of maps we have $h_0 = f\widetilde{f}$, while h_1 is the constant map. For the map $h(\cdot, s)$ the shaded area is mapped to p. This shows that every map has an inverse.

To see that the group $\pi_n(X, p))$, $n \geq 2$, is abelian, the reader is invited to look at the next figure, which shows a homotopy between fg and gf, where g and f are arbitrary spheroids. □

PROPOSITION 2.10.2. *For $n \geq 2$ and all path connected spaces X, the groups $\pi_n(X, p)$ and $\pi_n(X, q)$, where $p, q \in X$, are isomorphic, but the isomorphism is not canonical, it depends on the homotopy class (rel endpoints) of the path joining p to q.*

PROOF. The proof is similar to that of an analogous fact about the fundamental group. □

PROPOSITION 2.10.3. *The homotopy groups are homotopy invariants of path connected spaces.*

PROOF. The proof is a straightforward verification similar to that of an analogous statement about the fundamental group. □

EXERCISE 2.10.1. Prove that all the homotopy groups of a contractible space are trivial.

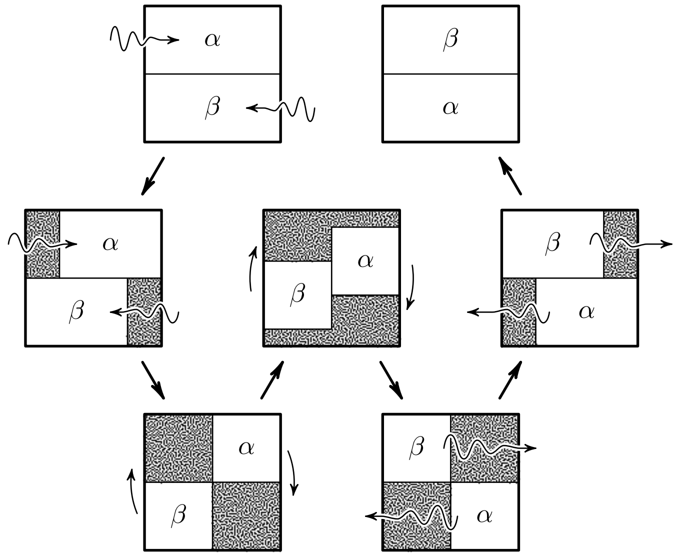

FIGURE 2.10.3. Multiplication of spheroids is commutative

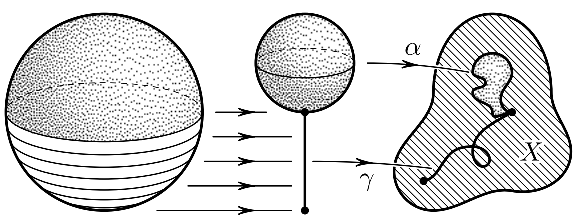

FIGURE 2.10.4. Change of basepoint isomorphism for spheroids

2.11. Hopf fibration

Unlike the fundamental group and homology groups (see Chapter 8), for which there exist general methods and algorithm for computation , the higher homotopy groups are extremely difficult to compute. The are certain "easy" cases: for example $\pi_k(\mathbb{S}^n) = 0$ for $k < n$ which we will be able to show by a proper extension of the method used in Proposition 2.8.14. Furthermore, $\pi_n(\mathbb{S}^n) = \mathbb{Z}$. this will be shown in **??** after developing a proper extension of degree theory.

However already computation of $\pi_k(\mathbb{S}^n) = 0$ for $k > n$ present very difficult problem which has not been completely solved. The first nontrivial example is the computation of π_3 for the sphere $\mathbb{S}^2$, based on one of the most beautiful constructions in topology – the Hopf fibration which we will describe now. Computation of $\pi_3(\mathbb{S}^2)$ is presented later in Chapter ??. The Hopf fibration appears in a number
ake more specific of problems in topology, geometry and differential equations.

Consider the unit sphere in $\mathbb{C}^2$:

$$\{(z_1, z_2) : |z_1|^2 + |z_2|^2 = 1\}$$

and the action H of the circle on it by scalar multiplication: for $\lambda \in \mathbb{S}^1$ put

$$H_\lambda(z_1, z_2) = (\lambda z_1, \lambda z_2).$$

PROPOSITION 2.11.1. *The identification space of this action is homeomorphic to* $\mathbb{S}^2$.

PROOF. This identification space is the same as the identification space of $\mathbb{C}^2$ where all proportional vectors are identified; it is simply the restriction of this equivalence relation to the unit sphere. The identification space is $\mathbb{C}P(1)$ which is homeomorphic to $\mathbb{S}^2$. □

The Hopf fibration is defined by a very simple formula. To help visualize we think of the sphere $\mathbb{S}^3$ as the one-point compactification of $\mathbb{R}^3$, so that we can actually draw the preimages of the Hopf map $h : \mathbb{S}^3 \to \mathbb{S}^2$ (which are circles) in the way shown on the figure.

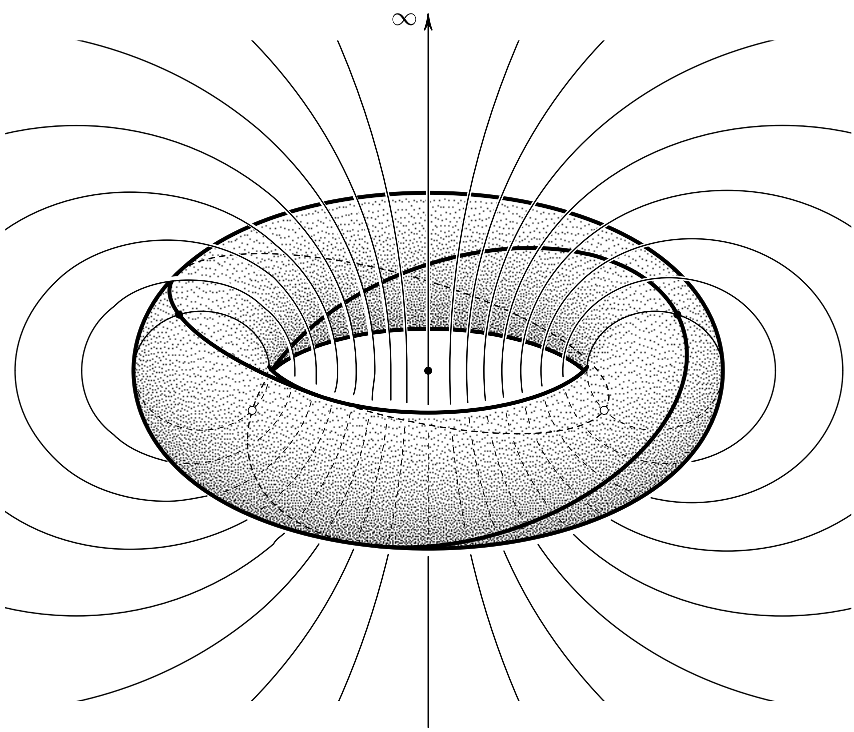

FIGURE 2.11.1. The Hopf fibration

EXERCISE 2.11.1. Let $z(t) = (z_1(t), z_2(t)) \in \mathbb{C}^2$. Consider the system of differential equations $\dot{z} = iz$ and restrict it to the 3-sphere $\{z \in \mathbb{C}^2 \mid |z| = 1\}$. Show that the trajectories of this system are circles constituting the Hopf fibration.

2.12. Problems

EXERCISE 2.12.1. Prove that in $\mathbb{S}^3$, represented as $\mathbb{R}^3 \cup \{\infty\}$, the complement of the unit circle in the xy-plane centered at the origin is homotopy equivalent to the circle.

EXERCISE 2.12.2. Prove that the 2-sphere with three points removed is homotopy equivalent to the figure eight (the wedge of two circles).

EXERCISE 2.12.3. The torus with three points removed is homotopy equivalent to the wedge of four circles.

EXERCISE 2.12.4. Let $f : \mathbb{S}^1 \to \mathbb{D}^2$ and $g : \mathbb{D}^2 \to \mathbb{S}^1$ be any continuous maps. Prove that their composition $g \circ f$ is homotopic to the constant map.

EXERCISE 2.12.5. For any finite cyclic group C there exists a compact connected three-dimensional manifold whose fundamental group is isomorphic to C.
Hint: Use the Hopf fibration.

EXERCISE 2.12.6. Show that the complex projective plane $\mathbb{C}P(2)$ (which is a four–dimensional manifold) is simply connected, i.e. its fundamental group is trivial.

EXERCISE 2.12.7. Consider the following map f of the torus $\mathbb{T}^2$ into itself:

$$f(x, y) = (x + \sin 2\pi y, 2y + x + 2\cos 2\pi x) \quad (\mod 1).$$

Describe the induced homomorphism f_* of the fundamental group.
Hint: You may use the description of the fundamental group of the direct product $\pi_1(X \times Y) = \pi_1(X) \times \pi_1(Y)$.

EXERCISE 2.12.8. Let $X = \mathbb{R}^2 \setminus \mathbb{Q}^2$. Prove that $\pi_1(X)$ is uncountable.

EXERCISE 2.12.9. The real projective space $\mathbb{R}P(n)$ is not simply connected.
Note: Use the fact that $\mathbb{R}P(n)$ is the sphere $\mathbb{S}^n$ with diametrically opposed points identified.

EXERCISE 2.12.10. For any abelian finitely generated group A there exists a compact manifold whose fundamental group is isomorphic to A.

EXERCISE 2.12.11. The fundamental group of any compact connected manifold is no more than countable and is finitely generated.

EXERCISE 2.12.12. Let X be the quotient space of the disjoint union of $\mathbb{S}^1$ and $\mathbb{S}^2$ with a pair of points $x \in S^1$ and $y \in S^2$ identified. Calculate $\pi_1(X)$.

CHAPTER 3

METRIC SPACES AND UNIFORM STRUCTURES

The general notion of topology does not allow to compare neighborhoods of different points. Such a comparison is quite natural in various geometric contexts. The general setting for such a comparison is that of a *uniform structure*. The most common and natural way for a uniform structure to appear is via a metric, which was already mentioned on several occasions in Chapter 1, so we will postpone discussing the general notion of union structure to Section 3.11 until after detailed exposition of metric spaces. Another important example of uniform structures is that of topological groups, see Section 3.12 below in this chapter. Also, as in turns out, a Hausdorff compact space carries a natural uniform structure, which in the separable case can be recovered from any metric generating the topology. Metric spaces and topological groups are the notions central for foundations of analysis.

3.1. Definition and basic constuctions

3.1.1. Axioms of metric spaces. We begin with listing the standard axioms of metric spaces, probably familiar to the reader from elementary real analysis courses, and mentioned in passing in Section 1.1, and then present some related definitions and derive some basic properties.

DEFINITION 3.1.1. If X is a set, then a function $d\colon X \times X \to \mathbb{R}$ is called a *metric* if

(1) $d(x, y) = d(y, x)$ *(symmetry)*,
(2) $d(x, y) \geq 0$; $d(x, y) = 0 \Leftrightarrow x = y$ *(positivity)*,
(3) $d(x, y) + d(y, z) \geq d(x, z)$ *(the triangle inequality)*.

If d is a metric, then (X, d) is called a *metric space*.

The set

$$B(x, r) := \{y \in X \mid d(x, y) < r\}$$

is called the *(open) r-ball* centered at x. The set

$$B_c(x, r) = \{y \in X \mid d(x, y) \leq r\}$$

is called the *closed r-ball* at (or around) x.

The *diameter* of a set in a metric space is the supremum of distances between its points; it is often denoted by diam A. The set A is called *bounded* if it has finite diameter.

A map $f : X \to Y$ between metric spaces with metrics d_X and d_Y is called as *isometric embedding* if for any pair of points $x, x' \in X$ $d_X(x, x') = d_Y(f(x), f(x'))$. If an isometric embedding is a bijection it is called an *isometry*. If there is an isometry between two metric spaces they are called *isometric*. This is an obvious equivalence relation in the category of metric spaces similar to homeomorphism for topological spaces or isomorphism for groups.

3.1.2. Metric topology. $O \subset X$ is called *open* if for every $x \in O$ there exists $r > 0$ such that $B(x, r) \subset O$. It follows immediately from the definition that open sets satisfy Definition 1.1.1. Topology thus defined is sometimes called the *metric topology* or *topology, generated by the metric d*. Naturally, different metrics may define the same topology.

Metric topology automatically has some good properties with respect to bases and separation.

Notice that the closed ball $B_c(x, r)$ contains the closure of the open ball $B(x, r)$ but may not coincide with it (Just consider the integers with the the standard metric: $d(m, n) = |m - n|$.)

Open balls as well as balls or rational radius or balls of radius $r_n,\ n = 1, 2, \dots,$ where r_n converges to zero, form a base of the metric topology.

PROPOSITION 3.1.2. *Every metric space is first countable. Every separable metric space has countable base.*

PROOF. Balls of rational radius around a point form a base of neighborhoods of that point.

By the triangle inequality, every open ball contains an open ball around a point of a dense set. Thus for a separable spaces balls of rational radius around points of a countable dense set form a base of the metric topology. □

Thus, for metric spaces the converse to Proposition 1.1.12 is also true.

Thus the closure of $A \subset X$ has the form

$$\bar{A} = \{x \in X \mid \forall r > 0, \quad B(x, r) \cap A \neq \varnothing\}.$$

For any closed set A and any point $x \in X$ the *distance from x to A*,

$$d(x, A) := \inf_{y \in A} d(x, y)$$

is defined. It is positive if and only if $x \in X \setminus A$.

THEOREM 3.1.3. *Any metric space is normal as a topological space.*

PROOF. For two disjoint closed sets $A, B \in X$, let

$$\mathcal{O}_A := \{x \in X \mid d(x, A) < d(x, B),\ \mathcal{O}_B := \{x \in X \mid d(x, B) < d(x, A).$$

These sets are open, disjoint, and contain A and B respectively. □

Let $\varphi : [0, \infty] \to \mathbb{R}$ be a nondecreasing, continuous, concave function such that $\varphi^{-1}(\{0\}) = \{0\}$. If (X, d) is a metric space, then $\phi \circ d$ is another metric on d which generates the same topology.

It is interesting to notice what happens if a function d as in Definition 3.1.1 does not satisfy symmetry or positivity. In the former case it can be symmetrized producing a metric $d_S(x, y) := \max(d(x, y), d(y, x))$. In the latter by the symmetry and triangle inequality the condition $d(x, y) = 0$ defines an equivalence relation and a genuine metric is defined in the space of equivalence classes. Note that some of the most impotrant notions in analysis such as spaces L^p of functions on a measure space are actually not spaces of actual functions but are such quotient spaces: their elements are equivalence classes of functions which coincide outside of a set of measure zero.

3.1.3. Constructions.

1. *Inducing*. Any subset A of a metric space X is a metric space with an *induced metric* d_A, the restriction of d to $A \times A$.

2. *Finite products*. For the product of finitely many metric spaces, there are various natural ways to introduce a metric. Let $\varphi : ([0, \infty])^n \to \mathbb{R}$ be a continuous concave function such that $\varphi^{-1}(\{0\}) = \{(0, \dots, 0)\}$ and which is nondecreasing in each variable.

Given metric spaces $(X_i, d_i),\ i = 1, \dots, n$, let

$$d^{\varphi} := \varphi(d_1, \dots, d_n) : (X_1 \times \dots X_n) \times (X_1 \times \dots X_n) \to \mathbb{R}.$$

EXERCISE 3.1.1. Prove that d^{φ} defines a metric on $X_1 \times \dots X_n$ which generates the product topology.

Here are examples which appear most often:

• the *maximum metric* corresponds to

$$\varphi(t_1, \dots, t_n) = \max(t_1, \dots, t_n);$$

• the l^p *metric* for $1 \le p < \infty$ corresponds to

$$\varphi(t_1, \dots, t_n) = (t_1^p + \dots + t_n^p)^{1/p}.$$

Two particularly important cases of the latter are $t = 1$ and $t = 2$; the latter produces the Euclidean metric in $\mathbb{R}^n$ from the standard (absolute value) metrics on n copies of $\mathbb{R}$.

3. *Countable products*. For a countable product of metric spaces, various metrics generating the product topology can also be introduced. One class of such metrics can be produced as follows. Let $\varphi : [0, \infty] \to \mathbb{R}$ be as above and let $a_1, a_2, \dots$ be a suquence of positive numbers such that the series $\sum_{n=1}^{\infty} a_n$ converges. Given metric spaces $(X_1, d_1), (X_2, d_2) \dots$, consider the metric d on the infinite product of the spaces $\{X_i\}$ defined as

$$d((x_1, x_2, \dots), (y_1, y_2, \dots)) := \sum_{n=1}^{\infty} a_n \varphi(d_n(x_n, y_n)).$$

EXERCISE 3.1.2. Prove that d is really a metric and that the corresponding metric topology coincides with the product topology.

4. *Factors*. On the other hand, projecting a metric even to a very good factor space is problematic. Let us begin with an example which exhibits some of the characteristic difficulties.

EXAMPLE 3.1.4. Consider the partition of the plane $\mathbb{R}^2$ into the level sets of the function xy, i.e. the hyperboli $xy = const \neq 0$ and the union of coordinate axes. The factor topology is nice and normal. It is easy to see in fact that the function xy on the factor space establishes a homeomorphism between this space and the real line. On the other hand, there is no natural way to define a metric in the factor space based on the Euclidean metric in the plane. Any two elements of the factor contain points arbitrary close to each other and arbitrary far away from each other so manipulating with infimums and supremums of of distances between the points in equivalence classes does not look hopeful.

We will see later that when the ambient space is compact and the factor-topology is Hausdorff there is a reasonable way to define a metric as the *Hausdorff metric* (see Definition 3.10.1) between equivalence classes considered as closed subsets of the space.

Here is a very simple but beautiful illustration how this may work.

EXAMPLE 3.1.5. Consider the real projective space $\mathbb{R}P(n)$ as the factor space of the sphere $\mathbb{S}^n$ with opposite points identified. Define the distance between the pairs $(x, -x)$ and $(y, -y)$ as the minimum of distances between members of the pairs. Notice that this minimum is achieved simultaneously on a pair and the pair of opposite points. This last fact allows to check the triangle inequality (positivity and symmetry are obvious) which in general would not be satisfied for the minimal distance of elements of equivalence classes even if those classes are finite.

EXERCISE 3.1.3. Prove the triangle inequality for this example. Prove that the natural projection from $\mathbb{S}^n$ to $\mathbb{R}P(n)$ is an isometric embedding in a neighborhood of each point. Calculate the maximal size of such a neighborhood.

Our next example is meant to demonstrate that the chief reason for the success of the previous example is not compactness but the fact that the factor space is the orbit space of an action by isometries (and of course is Hausdorff at the same time):

EXAMPLE 3.1.6. Consider the natural projection $\mathbb{R}^n \to \mathbb{R}^n/\mathbb{Z}^n = \mathbb{T}^n$. Define the distance $d(a\mathbb{Z}^n, b\mathbb{Z}^n)$ on the torus as the minimum of Euclidean distances between points in $\mathbb{R}^n$ in the equivalence classes representing corresponding points on the torus. Notice that since translations are isometries the minimum is always achieved and if it is achieved on a pair (x, y) it is also achieved on any integer translation of (x, y).

EXERCISE 3.1.4. Prove the triangle inequality for this example. Prove that the natural projection from $\mathbb{R}^n$ to $\mathbb{T}^n$ is an isometric embedding in any open ball of radius 1/2 and is not an isometric embedding in any open ball of any greater radius.

3.2. Cauchy sequences and completeness

3.2.1. Definition and basic properties. The notion of Cauchy sequence in Euclidean spaces and the role of its convergence should be familiar from elementary real analysis courses. Here we will review this notion in the most general setting, leading up to general theorems on completion, which play a crucial role in functional analysis.

DEFINITION 3.2.1. A sequence $\{x_i\}_{i\in\mathbb{N}}$ is called a *Cauchy sequence* if for all $\varepsilon > 0$ there exists an $N \in \mathbb{N}$ such that $d(x_i, x_j) < \varepsilon$ whenever $i, j \geq \mathbb{N}$; X is said to be *complete* if every Cauchy sequence converges.

PROPOSITION 3.2.2. *A subset A of a complete metric space X is a complete metric space with respect to the induced metric if and only if it is closed.*

PROOF. For a closed $A \in X$ the limit of any Cauchy sequence in A belongs to A. If A is not closed take a sequence in A converging to a point in $\bar{A} \setminus A$. It is Cauchy but does not converge in A. □

The following basic property of complete spaces is used in the next two theorems.

PROPOSITION 3.2.3. *Let $A_1 \supset A_2 \supset \dots$ be a nested sequence of closed sets in a complete metric space, such that* $\operatorname{diam} A_n \to 0$ *as $n \to \infty$. Then $\bigcap_{n=1}^{\infty} A_n$ is a single point.*

PROOF. Since $\operatorname{diam} A_n \to 0$ the intersection cannot contain more than one point. Take a sequence $x_n \in A_n$. It is Cauchy since $\operatorname{diam} A_n \to 0$. Its limit x belongs to $\bar{A}_n$ for any n. Since the sets A_i are closed, it follows that $x \in A_n$ for any n. □

3.2.2. The Baire category theorem.

THEOREM 3.2.4 (Baire Category Theorem). *In a complete metric space, a countable intersection of open dense sets is dense. The same holds for a locally compact Hausdorff space.*

PROOF. If $\{O_i\}_{i\in\mathbb{N}}$ are open and dense in X and $\varnothing \neq B_0 \subset X$ is open then inductively choose a ball B_{i+1} of radius at most ε/i for which we have $\bar{B}_{i+1} \subset O_{i+1} \cap B_i$. The centers converge by completeness, so

$$\varnothing \neq \bigcap_i \bar{B}_i \subset B_0 \cap \bigcap_i O_i.$$

For locally compact Hausdorff spaces take B_i open with compact closure and use the finite intersection property. □

The Baire Theorem motivates the following definition. If we want to mesure massivenes of sets in a topological or in particular metric space, we may assume that nowhere dense sets are small and their complements are massive. The next natural step is to introduce the following concept.

DEFINITION 3.2.5. Countable unions of nowhere dense sets are called *sets of first (Baire) category*.

The complement to a set of first baire category is called a *residual* set.

The Baire category theorem asserts that, at least for complete metric spaces, sets of first category can still be viewed as small, since they cannot fill any open set.

The Baire category theorem is a simple but powerful tool for proving *existence* of various objects when it is often difficult or impossible to produce those constructively.

3.2.3. Minimality of the Cantor set. Armed with the tools developed in the previous subsections, we can now return to the Cantor set and prove a universality theorem about this remarkable object.

THEOREM 3.2.6. *(cf. Exercise 1.10.14)*

Any uncountable separable complete metric space X contains a closed subset homeomorphic to the Cantor set.

PROOF. First consider the following subset

$$X_0\colon \{x \in X | \text{any neigbourhood of } x \text{ contains uncountably many points}\}$$

Notice that the set X_0 is perfect, i.e., it is closed and contains no isolated points.

LEMMA 3.2.7. *The set $X \setminus X_0$ is countable.*

PROOF. To prove the lemma, for each point $x \in X \setminus X_0$ find a neighborhood from a countable base which contains at most countably many points (Proposition 3.1.2). Thus $X \setminus X_0$ is covered by at most countably many sets each containing at most countably many points. □

Thus the theorem is a consequence of the following fact.

PROPOSITION 3.2.8. *Any perfect complete metric space X contains a closed subset homeomorphic to the Cantor set.*

PROOF. To prove the the proposition, pick two points $x_0 \neq x_1$ in X and let $d_0 := d(x_0, x_1)$. Let

$$X_i := \overline{B(x_i, (1/4)d_0)}, \quad i = 0, 1$$

and $C_1 := X_0 \cup X_1$.

Then pick two different points $x_{i,0}, x_{i,1} \in \operatorname{Int} X_i,\ i = 0,1$. Such choices are possible because any open set in X contains infinitely many points. Notice that $d(x_{i,0}, x_{i,1}) \le (1/2)d_0$. Let

$$Y_{i_1,i_2} := \overline{B(x_{i_1,i_2}, (1/4)d(x_{i_1,0}, x_{i_1,1}))},\ i_1, i_2 = 0,1,$$

$$X_{i_1,i_2} := Y_{i_1,i_2} \cap C_1 \quad \text{and } C_2 = X_{0,0} \cup X_{0,1} \cup X_{1,0} \cup X_{1,1}.$$

Notice that $\operatorname{diam}(X_{i_1,i_2}) \le d_0/2$.

Proceed by induction. Having constructed

$$C_n = \bigcup_{i_1,\dots,i_n \in \{0,1\}} X_{i_1,\dots,i_n}$$

with $\operatorname{diam} X_{i_1,\dots,i_n} \le d_0/2^n$, pick two different points $x_{i_1,\dots,i_n,0}$ and $x_{i_1,\dots,i_n,1}$ in $\operatorname{Int} X_{i_1,\dots,i_n}$ and let us successively define

$$Y_{i_1,\dots,i_n,i_{n+1}} := \bar{B}(x_{i_1,\dots,i_n,i_{n+1}}, d(x_{i_1,\dots,i_n,0}, x_{i_1,\dots,i_n,1})/4),$$

$$X_{i_1,\dots,i_n,i_{n+1}} := Y_{i_1,\dots,i_n,i_{n+1}} \cap C_n,$$

$$C_{n+1} := \bigcup_{i_1,\dots,i_n,i_{n+1} \in \{0,1\}} X_{i_1,\dots,i_n,i_{n+1}}.$$

Since $\operatorname{diam} X_{i_1,\dots,i_n} \le d_0/2^n$, each infinite intersection

$$\bigcap_{i_1,\dots,i_n,\dots \in \{0,1\}} X_{i_1,\dots,i_n,\dots}$$

is a single point by Heine–Borel (Proposition 3.2.3). The set $C := \bigcap_{n=1}^{\infty} C_n$ is homeomorphic to the countable product of the two point sets $\{0,1\}$ via the map

$$\bigcap_{i_1,\dots,i_n,\dots \in \{0,1\}} X_{i_1,\dots,i_n,\dots} \mapsto (i_1, \dots, i_n \dots).$$

By Proposition 1.7.3, C is homeomorphic to the Cantor set. □

The theorem is thus proved. □

3.2.4. Completion. Completeness allows to perform limit operations which arise frequently in various constructions. Notice that it is not possible to define the notion of Cauchy sequence in an arbitrary topological space, since one lacks the possibility of comparing neighborhoods at different points. Here the uniform structure (see Section 3.11) provides the most general natural setting.

A metric space can be made complete in the following way:

DEFINITION 3.2.9. If X is a metric space and there is an isometry from X onto a dense subset of a complete metric space $\hat{X}$ then $\hat{X}$ is called the *completion* of X.

THEOREM 3.2.10. *For any metric space X there exists a completion unique up to isometry which commutes with the embeddings of X into a completion.*

PROOF. The process mimics the construction of the real numbers as the completion of rationals, well–known from basic real analysis. Namely, the elements of the completion are equivalence classes of Cauchy sequences by identifying two sequences if the distance between the corresponding elements converges to zero. The distance between two (equivalence classes of) sequences is defined as the limit of the distances between the corresponding elements. An isometric embedding of X into the completion is given by identifying element of X with constant sequences. Uniqueness is obvious by definition, since by uniform continuity the isometric embedding of X to any completion extends to an isometric bijection of the standard completion. □

3.3. The p-adic completion of integers and rationals

This is an example which rivals the construction of real numbers in its importance for various areas of mathematics, especially to number theory and algebraic geometry. Unlike the construction of the reals, it gives infinitely many differnt nonisometric completions of the rationals.

3.3.1. The p-adic norm. Let p be a positive prime number. Any rational number r can be represented as $p^m\frac{k}{l}$ where m is an integer and k and l are integers realtively prime with p. Define the *p-adic norm* $\|r\|_p := p^{-m}$ and the distance $d_p(r_1, r_2) := \|r_1 - r_2\|_p$.

EXERCISE 3.3.1. Show that the p-adic norm is *multiplicative*, i.e., we have $\|r_1 \cdot r_2\|_p = \|r_1\|_p\|r_2\|_p$.

PROPOSITION 3.3.1. *The inequality*

$$d_p(r_1, r_3) \le \max(d_p(r_1, r_2), d_p(r_2, r_3))$$

holds for all $r_1, r_2, r_3 \in \mathbb{Q}$.

REMARK 3.3.2. A metric satisfying this property (which is stronger than the triangle inequality) is called an *ultrametric*.

PROOF. Since $\|r\|_p = \|-r\|_p$ the statement follows from the property of p-norms:

$$\|r_1 + r_2\|_p \le \|r_1\|_p + \|r_2\|_p.$$

To see this, write $r_i = p_i^m\frac{k_i}{l_i}$, $i = 1, 2$ with k_i and l_i relatively prime with p and assume without loss of generality that $m_2 \ge m_1$. We have

$$r_1 + r_2 = p_1^m\frac{k_1l_2 + p^{m_2-m_1}k_2l_1}{l_1l_2}.$$

The numerator $k_1l_2 + p^{m_2-m_1}k_2l_1$ is an integer and if $m_2 > m_1$ it is relatively prime with p. In any event we have $\|r_1+r_2\|_p \le p^{-m_1} = \|r_1\|_p = \max(\|r_1\|_p, \|r_2\|_p)$. □

3.3.2. The p-adic numbers and the Cantor set. Proposition 3.3.1 and the multiplicativity prorerty of the p-adic norm allow to extend addition and multiplication from $\mathbb{Q}$ to the completion. This is done in exacly the same way as in the real analysis for real numbers. The existence of the opposite and inverse (the latter for a nonzero element) follow easily.

Thus the completion becomes a field, which is called the *field of p-adic numbers* and is usually denoted by $\mathbb{Q}_p$. Restricting the procedure to the integers which always have norm ≤ 1 one obtains the subring of $\mathbb{Q}_p$, which is called the *ring of p-adic integers* and is usually denoted by $\mathbb{Z}_p$.

The topology of p–adic numbers once again indicates the importance of the Cantor set.

PROPOSITION 3.3.3. *The space $\mathbb{Z}_p$ is homeomorphic to the Cantor set; $\mathbb{Z}_p$ is the unit ball (both closed and open) in $\mathbb{Q}_p$.*

The space $\mathbb{Q}_p$ is homeomorphic to the disjoint countable union of Cantor sets.

PROOF. We begin with the integers. For any sequence

$$a = \{a_n\} \in \prod_{n=1}^{\infty}\{0, 1 \dots, p-1\}$$

the sequence of integers

$$k_n(a) := \sum_{i=1}^{n} a_n p^i$$

is Cauchy; for different $\{a_n\}$ these sequences are non equivalent and any Cauchy sequence is equivalent to one of these. Thus the correspondence

$$\prod_{n=1}^{\infty}\{0, 1 \dots, p-1\} \to \mathbb{Z}_p, \quad \{a_n\} \mapsto \text{the equivalence class of } k_n(a)$$

is a homeomorphism. The space $\prod_{n=1}^{\infty}\{0, 1 \dots, p-1\}$ can be mapped homeomorphically to a nowhere dense perfect subset of the interval by the map

$$\{a_n\}_{n=1}^{\infty} \mapsto \sum_{n=1}^{\infty} a_n (2p-1)^{-i}$$

. Thus the statement about $\mathbb{Z}_p$ follows from Proposition 1.7.5.

Since $\mathbb{Z}$ is the unit ball (open and closed) around 0 in the matric d_p and any other point is at a distance at least 1 from it, the same holds for the completions.

Finally, any rational number can be uniquely represented as

$$k + \sum_{i=1}^{n} a_i p^{-i}, \quad k \in \mathbb{Z}, \quad a_i \in \{0, \dots, p-1\},\ i = 1, \dots, n.$$

If the corresponging finite sequences a_i have different length or do not coincide, then the p-adic distance between the rationals is at least 1. Passing to the completion we see that any $x \in \mathbb{Q}_p$ is uniquely represented as $k + \sum_{i=1}^{n} a_i p^{-i}$ with $k \in \mathbb{Z}_p$. with pairwise distances for different a_i's at least one. $\square$

EXERCISE 3.3.2. Where in the construction is it important that p is a prime number?

3.4. Maps between metric spaces

3.4.1. Stronger continuity properties.

DEFINITION 3.4.1. A map $f : X \to Y$ between the metric spaces (X, d), (Y, dist) is said to be *uniformly continuous* if for all $\varepsilon > 0$ there is a $\delta > 0$ such that for all $x, y \in X$ with $d(x, y) < \delta$ we have $\mathrm{dist}(f(x), f(y)) < \varepsilon$. A uniformly continuous bijection with uniformly continuous inverse is called a *uniform homeomorphism*.

PROPOSITION 3.4.2. *A uniformly continuous map from a subset of a metric space to a complete space uniquely extends to its closure.*

PROOF. Let $A \subset X$, $x \in \bar{A}$, $f \colon A \to Y$ uniformly continuous. Fix an $\epsilon > 0$ and find the corresponding δ from the definition of uniform continuity. Take the closed $\delta/4$ ball around x. Its image and hence the closure of the image has diameter $\leq \epsilon$. Repeating this procedure for a sequence $\epsilon_n \to 0$ we obtain a nested sequence of closed sets whose diameters converge to zero. By Proposition 3.2.3 their intersection is a single point. If we denote this point by $f(x)$ the resulting map will be continuous at x and this extension is unique by uniqueness of the limit since by construction for any sequence $x_n \in A,\ x_n \to x$ one has $f(x_n) \to f(x)$. □

DEFINITION 3.4.3. A family $\mathcal{F}$ of maps $X \to Y$ is said to be *equicontinuous* if for every $x \in X$ and $\varepsilon > 0$ there is a $\delta > 0$ such that $d(x, y) < \delta$ implies

$$\mathrm{dist}(f(x), f(y)) < \varepsilon \quad \text{for all } y \in X \text{ and } f \in \mathcal{F}.$$

DEFINITION 3.4.4. A map $f \colon X \to Y$ is said to be *Hölder continuous* with exponent α, or α-*Hölder*, if there exist $C, \varepsilon > 0$ such that $d(x, y) < \varepsilon$ implies

$$d(f(x), f(y)) \leq C(d(x, y))^{\alpha},$$

Lipschitz continuous if it is 1-Hölder, and biLipschitz if it is Lipschitz and has a Lipschitz inverse.

It is useful to introduce local versions of the above notions. A map $f \colon X \to Y$ is said to be Hölder continuous with exponent α, at the point $x \in X$ or α-*Hölder*, if there exist $C, \varepsilon > 0$ such that $d(x, y) < \varepsilon$ implies

$$d(f(x), f(y)) \leq C(d(x, y))^{\alpha},$$

Lipschitz continuous at x if it is 1-Hölder at x.

3.4.2. Various equivalences of metric spaces. Besides the natural relation of isometry, the category of metric spaces is endowed with several other equivalence relations.

DEFINITION 3.4.5. Two metric spaces are *uniformly equivalent* if there exists a homeomorphism between the spaces which is uniformly continuous together with its inverse.

PROPOSITION 3.4.6. *Any metric space uniformly equivalent to a complete space is complete.*

PROOF. A uniformly continuous map obviously takes Cauchy sequences to Cauchy sequences. □

EXAMPLE 3.4.7. The open interval and the real line are homeomorphic but not uniformly equivalent because one is bounded and the other is not.

EXERCISE 3.4.1. Prove that an open half–line is not not uniformly equivalent to either whole line or an open interval.

DEFINITION 3.4.8. Metric spaces are *Hölder equivalent* if there there exists a homeomorphism between the spaces which is Hölder together with its inverse.

Metric spaces are *Lipschitz equivalent* if there exists a biLipschitz homeomorphism between the spaces.

EXAMPLE 3.4.9. Consider the standard middle–third Cantor set C and the subset C_1 of $[0, 1]$ obtained by a similar procedure but with taking away at every step the open interval in the middle of one half of the length. These two sets are Hólder equivalent but not Lipschitz equivalent.

EXERCISE 3.4.2. Find a Hölder homeomorphism with Hölder inverse in the previous example.

As usual, it is easier to prove existence of an equivalence that absence of one. For the latter one needs to produce an invariant of Lipschitz equivalence calculate it for two sets and show that the values (which do not have to be numbers but may be mathematical objects of another kind) are different. On this occasion one can use asymptotics of the minimal number of ϵ-balls needed to cover the set as $\epsilon \to 0$. Such notions are called *capacities* and are related to the important notion of *Hausdorff dimension* which, unlike the topological dimension, is not invariant under homeomorphisms. See **??**.

EXERCISE 3.4.3. Prove that the identity map of the product space is biLIpschitz homeomorphism between the space provided with the maximal metric and with any l^p metric.

EXAMPLE 3.4.10. The unit square (open or closed) is Lipschitz equivalent to the unit disc (respectively open or closed), but not isometric to it.

EXERCISE 3.4.4. Consider the unit circle with the metric induced from the $\mathbb{R}^2$ and the unit circle with the angular metric. Prove that these two metric spaces are Lipschitz equivalent but not isometric.

3.5. Role of metrics in geometry and topology

3.5.1. Elementary geometry. The study of metric spaces with a given metric belongs to the realm of geometry. The natural equivalence relation here is the strongest one, mentioned above, the isometry. Recall that the classical (or "elementary") Euclidean geometry deals with properties of simple objects in the plane or in the three-dimensional space invariant under isometries, or, according to some interpretations, under a larger class of similarity transformations since the absolute unit of length is not fixed in the Euclidean geometry (unlike the prototype non-Euclidean geometry, the hyperbolic one!).

Isometries tend to be rather rigid: recall that in the Euclidean plane an isometry is uniquely determined by images of three points (not on a line) , and in the Euclidean space by the images of four (not in a plane), and those images cannot be arbitrary.

EXERCISE 3.5.1. Prove that an isometry of $\mathbb{R}^n$ with the standard Euclidean metric is uniquely determined by images of any points $x_1, \dots, x_{n+1}$ such that the vectors $x_k - x_1,\ k = 2, \dots, n+1$ are linearly independent.

3.5.2. Riemannian geometry. The most important and most central for mathematics and physics generalization of Euclidean geometry is *Riemannian geometry*. Its objects are manifolds (in fact, differentiable or smooth manifolds which are defined and discussed in Chapter 4) with an extra structure of a *Riemannian metric* which defines Euclidean geometry (distances and angles) *infinitesimally* at each point, and the length of curves is obtained by integration. A smooth manifolds with a fixed Riemannian metric is called a *Riemannian manifold*. While we will wait till Section 13.2 for a systematic introduction to Riemannian geometry, instances of it have already appeared, e.g. the metric on the standard embedded sphere $\mathbb{S}^n \subset \mathbb{R}^{n+1}$ where the distance is measured along the great circles, (and is *not* induced from $\mathbb{R}^{n+1}$), its projection to $\mathbb{R}P(n)$, and projection of Euclidean metric in $\mathbb{R}^n$ to the torus $\mathbb{T}^n$. More general and more interesting classes of Riemannian manifolds will continue to pop up along the way, e.g. in **??** and **??**.

EXERCISE 3.5.2. Prove that in the spherical geometry the sum of angels of a triangle whose sides are arcs of great circles is always greater than π

3.5.3. More general metric geometries. Riemannian geometry is the richest and the most important but by no means only and not the most general way metric spaces appear in geometry. While Riemannian geometry, at least classically, has been inspired mostly by analytic methods of classical geometries (Euclidean, spherical and suchlike) there are other more contemporary directions which to a large extent are developing the synthetic methods of classical geometric reasoning; an outstanding example is the geometry of *Aleksandrov spaces*.

EXERCISE 3.5.3. Let $a > 0$ and denote by C_a the surface of the cone in $\mathbb{R}^3$ given by the conditions $a^2z^2 = x^2 + y^2$, $z \geq 0$. Call a curve in C_a a line segment if it is the shortest curve between its endpoints. Find all line segments in C_a.

3.5.4. Metric as a background and a base for Lipschitz structure. The most classical extensions of Euclidean geometry dealt (with the exception of spherical geometry) not with other metrics spaces but with geometric structures more general that Euclidean metric, such as affine and projective structures. To this one should add conformal structure which if of central importance for complex analysis. In all these geometries metrics appear in an auxiliary role such as the metric from Example 3.1.5 on real projective spaces.

EXERCISE 3.5.4. Prove that there is no metric on the projective line $\mathbb{R}P(1)$ generating the standard topology which is invariant under projective transformations.

EXERCISE 3.5.5. Prove that there is no metric in $\mathbb{R}^2$ generating the standard topology and invariant under all area preserving affine transformations, i.e transformations of the form $x \mapsto Ax + b$ where A is a matrix with determinanat ± 1 and b is a vector.

The role played by metrics in the principal branches of topology, algebraic and differential topology, is somewhat similar. Most spaces studied in those disciplines are metrizable; especially in the case of differential topology which studies smooth manifolds and various derivative objects, fixing a Riemannian metric on the manifold is very useful. It allows to bring precise measurements into the picture and provides various function spaces associated with the manifold such as spaces of smooth functions or differential forms, with the structure of a Banach space. But the choice of metric is usually arbitrary and only in the special cases, when the objects of study possess many symmetries, a particular choice of metric sheds much light on the core topological questions.

One should also point out that in the study of non-compact topological spaces and group actions on such spaces often a natural class of biLipschitz equivalent metrics appear. The study of such structures has gained importance over last two decades.

3.6. Separation properties and metrizability

As we have seen any metric topology is first countable (Proposition 3.1.2) and normal (Theorem 3.1.3). Conversely, it is natural to ask under what conditions a topological space has a metric space structure compatible with its topology.

A topological space is said to be *metrizable* if there exists a metric on it that induces the given topology. The following theorem gives necessary and sufficient conditions for metrizability for second countable topological spaces. Theorem 9.10 f

THEOREM 3.6.1. *[Urysohn Metrization Theorem]*
A normal space with a countable base for the topology is metrizable.

PROOF. ++++++++++++++++++++++++++ □

Theorem 3.6.1 and Proposition 1.5.4 imply

COROLLARY 3.6.2. *Any compact Hausdorff space with a countable base is metrizable.*

mple: normal first
le not metrizable?

3.7. Compact metric spaces

3.7.1. Sequential compactness.

PROPOSITION 3.7.1. *Any compact metric space is complete.*

PROOF. Suppose the opposite, that is, X is a compact metric space and a Cauchy sequence $x_n,\ n = 1, 2, \dots$ does not converge. By taking a subsebuence if necessary we may assume that all points x_n are different. The union of the elements of the sequence is closed since the sequence does not converge. Let

$$\mathcal{O}_n := X \setminus \bigcup_{i=n}^{\infty} \{x_n\}.$$

These sets form an open cover of X but since they are increasing there is no finite subcover. □

DEFINITION 3.7.2. Given $r > 0$ a subset A of a metric space X is called an *r-net* if for any $x \in X$ there is $a \in A$ such that the distance $d(x, a)$. Equivalently r-balls around the points of A cover X.

A set $A \subset X$ is called *r-separated* if the distance between any two different points in A is greater than r.

The following observation is very useful in the especially for quantifying the notion of compactness.

PROPOSITION 3.7.3. *Any maximal r-separated set is an r-net.*

PROOF. If A is r-separated and is not an r-net then there is a point $x \in X$ at a distance $\geq r$ from every point of A Hence the set $A \cup \{x\}$ is r-separated □

PROPOSITION 3.7.4. *The following properties of a metric space X are equivalent*

1. *X is compact;*
2. *for any $\epsilon > 0$ X contains a finite ϵ-net, or, equivalently, any r-separated set for any $r > 0$ is finite;*
3. *every sequence contains a congerving subsequence.*

PROOF. (1) → (2). If X is compact than the cover of X by all balls of radius ϵ contains a finite subcover; centers of those balls form a finite ϵ-net.

(2) → (3) By Proposition 3.7.1 it is sufficient to show that every sequence has a Cauchy subsequence. Take a sequence $x_n,\ n = 1, 2, \dots$ and consider a finite 1-net. There is a ball of radius 1 which contains infinitely many elements of the sequence. Consider only these elements as a subsequence. Take a finite 1/2-net and find a subsequence which lies in a single ball of radius 1/2. Continuing by induction we find nested subsequences of the original sequence which lie in balls of radius $1/2^n$. Using the standard diagonal process we construct a Cauchy subsequence.

(3) → (1). Let us first show that the space must be separable. This implies that any cover contains a countable subcover since the space has countable base. If the space is not separable than there exists an $\epsilon > 0$ such that for any countable (and hence finite) collection of points there is a point at the distance greater than ϵ from all of them. This allows to construct by induction an infinite sequence of points which are pairwise more than ϵ apart. Such a sequence obviously does not contain a converging subsequence.

Now assume there is an open countable cover $\{\mathcal{O}_1, \mathcal{O}_2, \dots\}$ without a finite subcover. Take the union of the first n elements of the cover and a point x_n outside of the union. The sequence $x_n,\ n = 1, 2, \dots$ thus defined has a converging subsequence $x_{n_k} \to x$. But x belong to a certain element of the cover, say $\mathcal{O}_N$. Then for a sufficinetly large k, $n_k > N$ hence $x_{n_k} \notin \mathcal{O}_N$, a contradiction to convergence. □

An immediate corollary of the proof is the following.

PROPOSITION 3.7.5. *Any compact metric space is separable.*

Aside from establishing equivalence of compactness and sequential compactness for metric spaces Proposition 3.7.4 contains a very useful criterion of compactness in the form of property (2). Right away it gives a necessary and sufficient condition for a (in general incomplete) metric space to have compact completion. As we see it later in Section 3.7.5 it is also a starting point for developing qualitative notions related to the "size" of a metric space.

DEFINITION 3.7.6. A metric space (X, d) is *totally bounded* if it contains a finite ϵ-net for any $\epsilon > 0$, or, equivalently if any r-separates subset of X for any $r > 0$ is finite.

Since both completion and any subset of a totally bounded space are totally bounded Proposition 3.7.4 immediately implies

COROLLARY 3.7.7. *Completion of a metric space is compact if and only if the space is totally bounded.*

EXERCISE 3.7.1. Prove that an isometric embedding of a compact metric space into itself is an isometry.

3.7.2. Lebesgue number.

PROPOSITION 3.7.8. *For an open cover of a compact metric space there exists a number δ such that every δ-ball is contained in an element of the cover.*

PROOF. Suppose the opposite. Then there exists a cover and a sequence of points x_n such that the ball $B(x_n, 1/2^n)$ does not belong to any element of the cover. Take a converging subsequence $x_{n_k} \to x$. Since the point x is covered by an open set, a ball of radius $r > 0$ around x belongs to that element. But for k large enough $d(x, x_{n_k}) < r/2$ and hence by the triangle inequality the ball $B(x_{n_k}, r/2)$ lies in the same element of the cover. □

The largest such number is called the *Lebesgue number* of the cover.

3.7.3. Characterization of Cantor sets.

THEOREM 3.7.9. *Any perfect compact totally disconnected metric space X is homeomorphic to the Cantor set.*

PROOF. Any point $x \in X$ is contained in a set of arbitrally small diameter which is both closed and open. For x is the intersection of all sets which are open and closed and contain x. Take a cover of $X \setminus X$ by sets which are closed and open and do not contain x Adding the ball $B(x, \epsilon)$ one obtains a cover of X which has a finite subcover. Union of elements of this subcover other than $B(x, \epsilon)$ is a set which is still open and closed and whose complement is contained in $B(x, \epsilon)$.

Now consider a cover of the space by sets of diameter ≤ 1 which are closed and open. Take a finite subcover. Since any finite intersection of such sets is still both closed an open by taking all possible intersection we obtain a *partition* of the space into finitely many closed and open sets of diameter ≤ 1. Since the space is perfect no element of this partition is a point so a further division is possible. Repeating this procedure for each set in the cover by covering it by sets of diameter $\leq 1/2$ we obtain a finer partition into closed and open sets of of diameter $\leq 1/2$. Proceeding by induction we obtain a nested sequence of finite partitions into closed and open sets of positive diameter $\leq 1/2^n$, $n = 0, 1, 2, \ldots$. Proceeding as in the proof of Proposition 1.7.5, that is, mapping elements of each partition inside a nested sequence of contracting intervals, we constuct a homeomorphism of the space onto a nowhere dense perfect subset of $[0, 1]$ and hence by Proposition 1.7.5 our space is homeomorphic to the Cantor set. □

3.7.4. Universality of the Hilbert cube. Theorem 3.2.6 means that Cantor set is in some sense a minimal nontrivial compact metrizable space. Now we will find a maximal one.

THEOREM 3.7.10. *Any compact separable metric space X is homeomorphic to a closed subset of the Hilbert cube H.*

PROOF. First by multiplying the metric by a constant if nesessary we may assume that the diameter of X is less that 1. Pick a dense sequence of points $x_1, x_2 \dots$ in X. Let $F : X \to H$ be defined by

$$F(x) = (d(x, x_1), d_(x, x_2), \dots).$$

This map is injective since for any two distict points x and x' one can find n such that $d(x, x_n) < (1/2)d(x', x_n)$ so that by the triangle inequality $d(x, x_n) < d(x', x_n)$ and hence $F(x) \neq F(x')$. By Proposition 1.5.11 $F(X) \subset H$ is compact and by Proposition 1.5.13 F is a homeomorphism between X and $F(X)$. □

EXERCISE 3.7.2. Prove that the infinite-dimensioanl torus $\mathbb{T}^\infty$, the product of the countably many copies of the unit circle, has the same universality property as the Hilbert cube, that is, any compact separable metric space X is homeomorphic to a closed subset of $\mathbb{T}^\infty$.

3.7.5. Capacity and box dimension. For a compact metric space there is a notion of the "size" or capacity inspired by the notion of volume. Suppose X is a compact space with metric d. Then a set $E \subset X$ is said to be *r-dense* if $X \subset \bigcup_{x \in E} B_d(x, r)$, where $B_d(x, r)$ is the r-ball with respect to d around x (see **??**). Define the *r-capacity* of (X, d) to be the minimal cardinality $S_d(r)$ of an r-dense set.

For example, if $X = [0, 1]$ with the usual metric, then $S_d(r)$ is approximately $1/2r$ because it takes over $1/2r$ balls (that is, intervals) to cover a unit length, and the $\lfloor 2 + 1/2r \rfloor$-balls centered at $ir(2 - r)$, $0 \leq i \leq \lfloor 1 + 1/2r \rfloor$ suffice. As another example, if $X = [0, 1]^2$ is the unit square, then $S_d(r)$ is roughly r^{-2} because it takes at least $1/\pi r^2$ r-balls to cover a unit area, and, on the other hand, the $(1 + 1/r)^2$-balls centered at points (ir, jr) provide a cover. Likewise, for the unit cube $(1 + 1/r)^3$, r-balls suffice.

In the case of the ternary Cantor set with the usual metric we have $S_d(3^{-i}) = 2^i$ if we cheat a little and use closed balls for simplicity; otherwise, we could use $S_d((3 - 1/i)^{-i}) = 2^i$ with honest open balls.

One interesting aspect of capacity is the relation between its dependence on r [that is, with which power of r the capacity $S_d(r)$ increases] and dimension.

If $X = [0, 1]$, then

$$\lim_{r \to 0} -\frac{\log S_d(r)}{\log r} \geq \lim_{r \to 0} -\frac{\log(1/2r)}{\log r} = \lim_{r \to 0} \frac{\log 2 + \log r}{\log r} = 1$$

and

$$\lim_{r \to 0} -\frac{\log S_d(r)}{\log r} \leq \lim_{r \to 0} -\frac{\log\lfloor 2 + 1/2r \rfloor}{\log r} \leq \lim -\frac{\log(1/r)}{\log r} = 1,$$

so $\lim_{r \to 0} -\log S_d(r)/\log r = 1 = \dim X$. If $X = [0, 1]^2$, then

$$\lim_{r \to 0} -\log S_d(r)/\log r = 2 = \dim X,$$

and if $X = [0,1]^3$, then

$$\lim_{r\to 0} -\log S_d(r)/\log r = 3 = \dim X.$$

This suggests that $\lim_{r\to 0} -\log S_d(r)/\log r$ defines a notion of dimension.

DEFINITION 3.7.11. If X is a totally bounded metric space (Definition 3.7.6), then

$$\operatorname{bdim}(X) := \lim_{r\to 0} -\frac{\log S_d(r)}{\log r}$$

is called the *box dimension* of X.

Let us test this notion on a less straightforward example. If C is the ternary Cantor set, then

$$\operatorname{bdim}(C) = \lim_{r\to 0} -\frac{\log S_d(r)}{\log r} = \lim_{n\to\infty} -\frac{\log 2^i}{\log 3^{-i}} = \frac{\log 2}{\log 3}.$$

If C_α is constructed by deleting a middle interval of relative length $1-(2/\alpha)$ at each stage, then $\operatorname{bdim}(C_\alpha) = \log 2/\log\alpha$. This increases to 1 as $\alpha \to 2$ (deleting ever smaller intervals), and it decreases to 0 as $\alpha \to \infty$ (deleting ever larger intervals). Thus we get a small box dimension if in the Cantor construction the size of the remaining intervals decreases rapidly with each iteration.

This illustrates, by the way, that the box dimension of a set may change under a homeomorphism, because these Cantor sets are pairwise homeomorphic. Box dimension and an associated but more subtle notion of *Hausdorff dimension* are the prime exhibits in the panoply of "fractal dimensions", the notion surrounded by a certain mystery (or mystique) at least for laymen. In the next section we will present simple calculations which shed light on this notion.

3.8. Metric spaces with symmetries and self-similarities

3.8.1. Euclidean space as an ideal geometric object and some of its close relatives. An outstanding, one may even say, the central, feature of Euclidean geometry, is an abundance of isometries in the Euclidean space. Not only there is isometry which maps any given point to any other point (e.g. the parallel translation by the vector connecting those points) but there are also isometries which interchange any given pair of points, e.g the central symmetry with respect to the midpoint of the interval connecting those points, or the reflection in the (hyper)plane perpendicular to that interval at the midpoint. The latter property distinguishes a very important class of Riemannian manifolds, called *symmetric spaces*. The next obvious examples of symmetric space after the Euclidean spaces are spheres $\mathbb{S}^n$ with the standard metric where the distance is measure along the shorter arcs of great circles. Notice that the metric induced from the embedding of $\mathbb{S}^n$ as the unit sphere into $\mathbb{R}^{n+1}$ also possesses all there isometries but the metric is not a Riemanninan metric, i.e. the distance cannot be calculated as the minimum of lengths of curves connecting two points, and thus this metric is much less interesting.

EXERCISE 3.8.1. How many isometries are there that interchange two points $x, y \in \mathbb{R}^n$ for different values of n?

EXERCISE 3.8.2. How many isometries are there that interchange two points $x, y \in \mathbb{S}^n$ for different values of n and for different configurations of points?

EXERCISE 3.8.3. Prove that the real projective space $\mathbb{R}P(n)$ with the metric inherited from the sphere (**??**) is a symmetric space.

EXERCISE 3.8.4. Prove that the torus $\mathbb{T}^n$ is with the metric inherited from $\mathbb{R}^n$ a symmetric space.

There is yet another remarkable property of Euclidean spaces which is not shared by other symmetric spaces: existence of *similarities*, i.e. transformations which preserve angles and changes all distances with the same coefficient of proportionality. It is interesting to point out that in the long quest to "prove" Euclid's fifth postulate, i.e. to deduce it from other axioms of Euclidean geometry, one among many equivalent formulations of the famous postulate is existence of a single pair of similar but not equal (not isometric) triangles. In the non-Euclidean hyperbolic geometry which results from adding the negation of the fifth postulates there no similar triangles and instead there is absolute unit of length! Incidentally the hyperbolic plane (as well as its higher-dimensional counterparts) is also a symmetric space. Existence of required symmetries can be deduced synthetically form the axioms common to Euclidean and non-Euclidean geometry, i.e. it belong s to so-called *absolute geometry*, the body of statement which can be proven in Euclidean geometry without the use of fifth postulate.

Metric spaces for which there exists a self-map which changes all distance with the same coefficient of proportionality different from one are called *self-similar*.

Obviously in a compact globally self-similar space which contain more one point the coefficient of proportionality for any similarity transformation must be less than one and such a transformation cannot be bijective; for non-compact spaces this is possible however.

3.8.2. Metrics on the Cantor set with symmetries and self-similarities. There is an interesting example of a similarity on the middle-third Cantor set, namely, $f_0\colon [0,1] \to [0,1]$, $f_0(x) = x/3$. Since f_0 is a contraction, it is also a contraction on every invariant subset, and in particular on the Cantor set. The unique fixed point is obviously 0. There is another contraction with the same contraction coefficient 1/3 preserving the Cantor set, namely $f_1(x) = \frac{x+2}{3}$ with fixed point 1. Images of these two contractions are disjoint and together they cover the whole Cantor set

EXERCISE 3.8.5. Prove that any similarity of the middle third Cantor set belongs to the semigroup generated by f_0 and f_1.

EXERCISE 3.8.6. Find infinitely many different self-similar Cantor sets on $[0,1]$ which contain both endpoints 0 and 1.

FIGURE 3.8.1. Sierpinski carpet and Sierpinski gasket.

FIGURE 3.8.2. The Koch snowflake.

3.8.3. Other Self-Similar Sets. Let us describe some other interesting self-similar metric spaces that are of a different form. The *Sierpinski carpet* (see **??**) is obtained from the unit square by removing the "middle-ninth" square $(1/3, 2/3) \times (1/3, 2/3)$, then removing from each square $(i/3, i + 1/3) \times (j/3, j + 1/3)$ its "middle ninth," and so on. This construction can easily be described in terms of ternary expansion in a way that immediately suggests higher-dimensional analogs.

Another very symmetric construction begins with an equilateral triangle with the bottom side horizontal, say, and divide it into four congruent equilateral triangles of which the central one has a horizontal top side. Then one deletes this central triangle and continues this construction on the remaining three triangles. he resulting set is sometimes called *Sierpinski gasket*.

The *von Koch snowflake* is obtained from an equilateral triangle by erecting on each side an equilateral triangle whose base is the middle third of that side and continuing this process iteratively with the sides of the resulting polygon It is attributed to Helge von Koch (1904).

A three-dimensional variant of the Sierpinski carpet S is the Sierpinski sponge or Menger curve defined by $\{(x, y, z) \in [0, 1]^3 \mid (x, y) \in S,\ (x, z) \in S\ (y, z) \in S\}$. It is obtained from the solid unit cube by punching a 1/3-square hole through the center from each direction, then punching, in each coordinate direction, eight 1/9-square holes through in the right places, and so on. Both Sierpinski carper and Menger curve have important universality properties which we do not discuss in this book.

Let as calculate the box dimension of these new examples. For the square Sierpinski carpet we can cheat as in the capacity calculation for the ternary Cantor set and use closed balls (sharing their center with one of the small remaining cubes at a certain stage) for covers. Then $S_d(3^{-i}/\sqrt{2}) = 8^i$ and

$$\mathrm{bdim}(S) = \lim_{n\to\infty} -\frac{\log 8^i}{\log 3^{-i}/\sqrt{2}} = \frac{\log 8}{\log 3} = \frac{3\log 2}{\log 3},$$

which is three times that of the ternary Cantor set (but still less than 2, of course). For the triangular Sierpinski gasket we similarly get box dimension $\log 3/\log 2$.

The Koch snowflake K has $S_d(3^{-i}) = 4^i$ by covering it with (closed) balls centered at the edges of the ith polygon. Thus

$$\mathrm{bdim}(K) = \lim_{n\to\infty} -\frac{\log 4^i}{\log 3^{-i}} = \frac{\log 4}{\log 3} = \frac{2\log 2}{\log 3},$$

which is less than that of the Sierpinski carpet, corresponding to the fact that the iterates look much "thinner". Notice that this dimension exceeds 1, however, so it is larger than the dimension of a curve. All of these examples have (box) dimension that is not an integer, that is, fractional or "fractal". This has motivated calling such sets *fractals*.

Notice a transparent connection between the box dimension and coefficients of self-similarity on all self-similar examples.

3.9. Spaces of continuous maps

If X is a compact metrizable topological space (for example, a compact manifold), then the space $C(X, X)$ of continuous maps of X into itself possesses the C^0 or *uniform* topology. It arises by fixing a metric ρ in X and defining the distance d between $f, g \in C(X, X)$ by

$$d(f, g) := \max_{x \in X} \rho(f(x), g(x)).$$

The subset $\mathrm{Hom}(X)$ of $C(X, X)$ of homeomorphisms of X is neither open nor closed in the C^0 topology. It possesses, however, a natural topology as a complete metric space induced by the metric

$$d_H(f, g) := \max(d(f, g), d(f^{-1}, g^{-1})).$$

If X is σ-compact we introduce the compact–open topologies for maps and homeomorphisms, that is, the topologies of uniform convergence on compact sets.

We sometimes use the fact that equicontinuity gives some compactness of a family of continuous functions in the uniform topology.

THEOREM 3.9.1 (Arzelá–Ascoli Theorem). *Let X, Y be metric spaces, X separable, and $\mathcal{F}$ an equicontinuous family of maps. If $\{f_i\}_{i \in \mathbb{N}} \subset \mathcal{F}$ such that $\{f_i(x)\}_{i \in \mathbb{N}}$ has compact closure for every $x \in X$ then there is a subsequence converging uniformly on compact sets to a function f.*

Thus in particular a closed bounded equicontinuous family of maps on a compact space is compact in the uniform topology (induced by the maximum norm).

Let us sketch the proof. First use the fact that $\{f_i(x)\}_{i \in \mathbb{N}}$ has compact closure for every point x of a countable dense subset S of X. A diagonal argument shows that there is a subsequence f_{i_k} which converges at every point of S. Now equicontinuity can be used to show that for every point $x \in X$ the sequence $f_{i_k}(x)$ is Cauchy, hence convergent (since $\{f_i(x)\}_{i \in \mathbb{N}}$ has compact, hence complete, closure). Using equicontinuity again yields continuity of the pointwise limit. Finally a pointwise convergent equicontinuous sequence converges uniformly on compact sets.

EXERCISE 3.9.1. Prove that the set of Lipschitz real-valued functions on a compact metric space X with a fixed Lipschitz constant and bounded in absolute value by another constant is compact in $C(x, \mathbb{R})$.

EXERCISE 3.9.2. Is the closure in $C([0, 1], \mathbb{R})$ (which is usually denoted simply by $C([0, 1])$) of the set of all differentiable functions which derivative bounded by 1 in absolute value and taking value 0 at 1/2 compact?

3.10. Spaces of closed subsets of a compact metric space

3.10.1. Hausdorff distance: definition and compactness. An interesting construction in the theory of compact metric spaces is that of the Hausdorff metric:

DEFINITION 3.10.1. If (X, d) is a compact metric space and $K(X)$ denotes the collection of closed subsets of X, then the *Hausdorff metric* d_H on $K(X)$ is defined by

$$d_H(A, B) := \sup_{a \in A} d(a, B) + \sup_{b \in B} d(b, A),$$

where $d(x, Y) := \inf_{y \in Y} d(x, y)$ for $Y \subset X$.

Notice that d_H is symmetric by construction and is zero if and only if the two sets coincide (here we use that these sets are closed, and hence compact, so the "sup" are actually "max"). Checking the triangle inequality requires a little extra work. To show that $d_H(A, B) \leq d_H(A, C) + d_H(C, B)$, note that $d(a, b) \leq d(a, c) + d(c, b)$ for $a \in A$, $b \in B$, $c \in C$, so taking the infimum over b we get $d(a, B) \leq d(a, c) + d(c, B)$ for $a \in A$, $c \in C$. Therefore, $d(a, B) \leq d(a, C) + \sup_{c \in C} d(c, B)$ and $\sup_{a \in A} d(a, B) \leq \sup_{a \in A} d(a, C) + \sup_{c \in C} d(c, B)$. Likewise, one gets $\sup_{b \in B} d(b, A) \leq \sup_{b \in B} d(b, C) + \sup_{c \in C} d(c, A)$. Adding the last two inequalities gives the triangle inequality.

PROPOSITION 3.10.2. *The Hausdorff metric on the closed subsets of a compact metric space defines a compact topology.*

PROOF. We need to verify total boundedness and completeness. Pick a finite $\epsilon/2$-net N. Any closed set $A \subset X$ is covered by a union of ϵ-balls centered at points of N, and the closure of the union of these has Hausdorff distance at most ϵ from A. Since there are only finitely many such sets, we have shown that this metric is totally bounded. To show that it is complete, consider a Cauchy sequence (with respect to the Hausdorff metric) of closed sets $A_n \subset X$. If we let $A := \bigcap_{k \in \mathbb{N}} \overline{\bigcup_{n \geq k} A_n}$, then one can easily check that $d(A_n, A) \to 0$. □

EXERCISE 3.10.1. Prove that for the Cantor set C the space $K(C)$ is homeomorphic to C.

EXERCISE 3.10.2. Prove that $K([0, 1])$ contains a subset homeomorphic to the Hilbert cube.

3.10.2. Existence of a minimal set for a continuous map. Any homeomorphism of a compact metric space X induces a natural homeomorphism of the collection of closed subsets of X with the Hausdorff metric, so we have the following:

PROPOSITION 3.10.3. *The set of closed invariant sets of a homeomorphism f of a compact metric space is a closed set with respect to the Hausdorff metric.*

PROOF. This is just the set of fixed points of the induced homeomorphism; hence it is closed. □

We will now give a nice application of the Hausdorff metric. Brouwer fixed point Theorem (Theorem 2.5.1 and Theorem 9.3.7) does not extend to continuous maps of even very nice spaces other than the disc. The simplest example of a continuous map (in fact a self–homeomorphism) which does not have have fixed points is a rotation of the circle; if the angle of rotation is a rational multiple of π all points are periodic with the same period; otherwise there are no periodic points. However, there is a nice generalization which works for any compact Hausdorff spaces. An obvious property of a fixed or periodic point for a continuous map is its minimality: it is an invariant closed set which has no invariant subsets.

DEFINITION 3.10.4. An invariant closed subset A of a continuous map $f\colon X \to X$ is *minimal* if there are no nonempty closed f-invariant subsets of A.

THEOREM 3.10.5. *Any continuous map f of a compact Hausdorff space X with a countable base into itself has an invariant minimal set.*

PROOF. By Corollary 3.6.2 the space X is metrizable. Fix a metric d on X and consider the Hausdorff metric on the space $K(X)$ of all closed subsets of X. Since any closed subset A of X is compact (Proposition 1.5.2) $f(A)$ is also compact (Proposition 1.5.11) and hence closed (Corollary 3.6.2). Thus f naturally induces a map $f_*\colon K(X) \to K(X)$ by setting $f_*(A) = f(A)$. A direct calculation shows that the map f_* is continuous in the topology induced by the Hausdorff metric. Closed f-invariant subsets of X are fixed points of f_*. The set of all such sets is closed, hence compact subset $I(f)$ of $K(X)$. Consider for each $B \in I(f)$ all $A \in I(f)$ such that $A \subset B$. Such A form a closed, hence compact, subset $I_B(f)$. Hence the function on $I_B(f)$ defined by $d_H(A, B)$ reaches its maximum, which we denote by $m(B)$, on a certain f-invariant set $M \subset B$.

Notice that the function $m(B)$ is also continuous in the topology of Hausdorff metric. Hence it reaches its minimum m_0 on a certain set N. If $m_0 = 0$, the set N is a minimal set. Now assume that $m_0 > 0$.

Take the set $M \subset B$ such that $d_H(M, B) = m(B) \geq m_0$. Inside M one can find an invariant subset M_1 such that $d_H(M_1, M) \geq m_0$. Notice that since $M_1 \subset M$, $d_H(M_1, B) \geq d_H(M, B) = m(B) \geq m_0$.

Continuing by induction we obtain an infinite sequence of nested closed invariant sets $B \supset M \supset M_1 \supset M_2 \supset \cdots \supset M_n \supset \dots$ such that the Hausdorff distance between any two of those sets is at least m_0. This contradicts compactness of $K(X)$ in the topology generated by the Hausdorff metric. $\square$

EXERCISE 3.10.3. Give detailed proofs of the claims used in the proof of Theorem 3.10.5:

- the map $f_*\colon K(X) \to K(X)$ is continuous;
- the function $m(\cdot)$ is continuous;
- $d_H(M_i, M_j) \geq m_0$ for $i, j = 1, 2, \dots; i \neq j$.

EXERCISE 3.10.4. For every natural number n give an example of a homeomorphism of a compact path connected topological space which has no fixed points and has exactly n minimal sets.

3.11. Uniform structures

3.11.1. Definitions and basic properties. The main difference between a metric topology and an even otherwise very good topology defined abstractly is the possibility to choose "small" neighborhoods for all points in the space simultaneously; we mean of course fixing an (arbitrary small) positive number r and taking balls $B(x, r)$ for all x. The notion of uniform structure is a formalization of such a possibility without metric (which is not always possible under the axioms below)

3.11.2. Uniform structure associated with compact topology.

3.12. Topological groups

In this section we introduce groups which carry a topology invariant under the group operations. A *topological group* is a group endowed with a topology with respect to which all *left translations* $L_{g_0}\colon g \mapsto g_0 g$ and *right translations* $R_{g_0}\colon g \mapsto g g_0$ as well as $g \mapsto g^{-1}$ are homeomorphisms. Familiar examples are $\mathbb{R}^n$ with the additive structure as well as the circle or, more generally, the n-torus, where translations are clearly diffeomorphisms, as is $x \mapsto -x$.

3.13. Problems

EXERCISE 3.13.1. Prove that every metric space is homeomorphic to a bounded space.

EXERCISE 3.13.2. Prove that in a compact set A in metric space X there exists a pair or points $x, y \in A$ such that $d(x, y) = \operatorname{diam} A$.

EXERCISE 3.13.3. Suppose a function $d\colon X \times X \to \mathbb{R}$ satisfies conditions (2) and (3) of Definition 3.1.1 but not (1). Find a natural way to modify this function so that the modified function becomes a metric.

EXERCISE 3.13.4. Let S be a smooth surface in $\mathbb{R}^3$, i.e. it may be a non-critical level of a smooth real-valued function, or a closed subset locally given as a graph when one coordinate is a smooth function of two others. S carries two metrics: (i) induced from $\mathbb{R}^3$ as a subset of a metric space, and (ii) the natural internal distance given by the minimal length of curves in S connecting two points.

Prove that if these two metrics coincide then S is a plane.

EXERCISE 3.13.5. Introduce a metric d on the Cantor set C (generating the Cantor set topology) such that (C, d) cannot be isometrically embedded to $\mathbb{R}^n$ for any n.

EXERCISE 3.13.6. Introduce a metric d on the Cantor set C such that (C, d) is not Lipschitz equivalent to a subset of $\mathbb{R}^n$ for any n.

EXERCISE 3.13.7. Prove that the set of functions which are not Hölder continuous at any point is a residual subset of $C([0, 1])$.

EXERCISE 3.13.8. Let $f\colon [0, 1]\mathbb{R}^2$ be α-Höder with $\alpha > 1/2$. Prove that $f([0, 1)]$ is nowhere dense.

EXERCISE 3.13.9. Find a generalization of the previous statement for the maps of the m-dimensional cube I^m to $\mathbb{R}^n$ with $m < n$.

EXERCISE 3.13.10. Prove existence of 1/2-Hölder surjective map $f\colon [0, 1] \to I^2$. (Such a map is usually called a *Peano curve*).

EXERCISE 3.13.11. Prove that any connected topological manifold is metrizable.

EXERCISE 3.13.12. Find a Riemannian metric on the complex projective space $\mathbb{C}P(n)$ which makes it a symmetric space.

EXERCISE 3.13.13. Prove that $\mathbb{S}^n$ is not self-similar.

CHAPTER 4

REAL AND COMPLEX SMOOTH MANIFOLDS

The notion of smooth or differentiable manifold is one of the central concepts of modern mathematics and its applications, and is also of fundamental importance in theoretical mechanics and physics. Roughly speaking, a smooth manifold is a topological space which may have a complicated global structure, but locally is like Euclidean space, i.e. it is a topological manifold as in Section 1.8 (it possesses "local coordinates"), with the transition from one system of local coordinates to a neighboring one being ensured by smooth functions. The fact that the transition functions are smooth allows the use of the whole machinery of the multivariable differential and integral calculus, which interacts very efficiently with geometric and topological tools in that setting.

This chapter is only a first introduction to real (and complex) smooth manifolds. We will return to this topic in Chapter 10, where, after having further developed some of these tools, in particular homology theory, we will have a glance at deep connections between algebraic and differential topology.

4.1. Differentiable manifolds, smooth maps and diffeomorphisms

4.1.1. Definitions.

DEFINITION 4.1.1. A Hausdorff topological space M with countable base is said to be an *n-dimensional differentiable* (or *smooth*) *manifold* if it is covered by a family $\mathcal{A} = \{(U_\alpha, h_\alpha)\}_{\alpha\in A}$ of open sets U_α called *charts* and supplied with homeomorphisms into $\mathbb{R}^n$,

$$\bigcup_\alpha U_\alpha = M, \quad h_\alpha : U_\alpha \to \mathbb{R}^n$$

(the index set A may be finite, countable or uncountable) that satisfy the *compatibility* condition: for any two charts (U_1, h_1) and (U_2, h_2) in $\mathcal{A}$ with $h_i \colon U_i \to B_i \subset \mathbb{R}^n$ the *coordinate change* $h_2 \circ h_1^{-1}$ (also sometimes called *transition function*) is differentiable on $h_1(U_1 \cap U_2) \subset B_1$.

Here "differentiable" can be taken to mean C^r for any $r \in \mathbb{N} \cup \infty$, or analytic. A collection of such charts covering M is called an *atlas* of M. Any atlas defines a unique *maximal atlas* obtained by taking all charts compatible with the present ones. A maximal atlas is called a *differentiable* (or *smooth*) *structure*.

DEFINITION 4.1.2. A *smooth* or *differentiable map* of one smooth manifold to another is a map $f : M \to N$ which is expressed by differentiable functions in the local coordinates of any chart. More precisely, for any charts (U, h), $U \ni x$

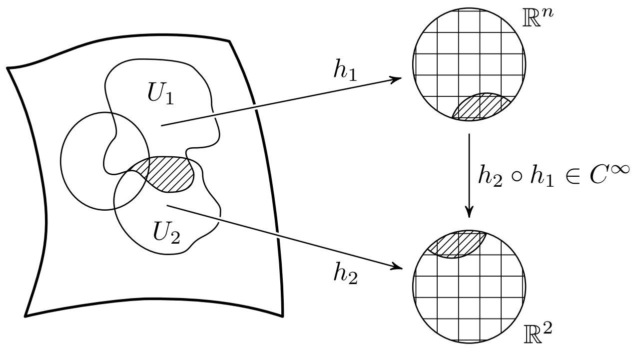

FIGURE 4.1.1. Definition of a smooth manifold

and (V, k), $V \ni f(x)$, the map $k \circ f \circ h^{-1}$ is a differentiable map of one domain of Euclidean space into another.

In view of the compatibility condition, in order to check the smoothness of a map $f : M \to N$ it suffices to check that it is smooth on any cover of M by charts and not on all charts of the maximal atlas.

DEFINITION 4.1.3. A *diffeomorphism* between smooth manifolds is a bijective smooth map with smooth inverse.

Obviously, any diffeomorphism is a homeomorphism,

EXERCISE 4.1.1. Give an example of a homeomorphism which is not a diffeomorphism.

The notion of diffeomorphism provides the natural concept of isomorphism of the smooth structures of manifolds: diffeomorphic manifolds are undistinguishable as differentiable manifolds. [1]

REMARK 4.1.4. In the definition above the local model for a differentiable manifold is $\mathbb{R}^n$ with its differentiable structure. It follows form the definition that any manifold diffeomorphic to $\mathbb{R}^n$ may serve as an alternative model. Two useful special cases are an open ball in $\mathbb{R}^n$ and the open unit disc $(0, 1)^n$. This follows from the fact that all those models are diffeomorphic smooth manifolds, see Exercise 4.1.2 and Exercise 4.1.3.

The Inverse Function Theorem from multi-variable calculus provides the following criterion which can be checked on most occasions.

PROPOSITION 4.1.5. *A map $f\colon M \to N$ between differentiable manifolds is a diffeomorphism if and only if (i) it is bijective and (ii) there exists atlases $\mathcal{A}$ and $\mathcal{B}$ for the differentiable structures in M and N correspondingly such that for any $x \in M$ there exist $(U, h) \in \mathcal{A},\ x \in U$ with $h(x) = p \in \mathbb{R}^n$ and $(V, k) \in \mathcal{B},\ f(x) \in V$ such that* $\det A \neq 0$ *where A the matrix of partial derivatives of $h^{-1} \circ f \circ k$ at p.*

[1]However, the same manifold may have different representations which, for example, may carry different geometric structures.

PROOF. Necessity follows directly from the definition.

To prove sufficiency first notice that by the chain rule for the coordinate changes in $\mathbb{R}^n$ condition (ii) is independent from a choice of (U, h) and (V, k) from the maximal atlases providing $x \in U$ and $f(x) \in V$.

The Inverse Function Theorem guarantees that $h^{-1} \circ f \circ k$ is a diffeomorphism between a sufficiently small ball around p and its image. Taking such balls and their images for covers of M and N correspondingly we see that both f and f^{-1} are smooth. □

PROPOSITION 4.1.6. *Let M be a differentiable manifold, $A \subset M$ an open set. Then A has a natural structure of differentiable manifold compatible with that for M.*

PROOF. Let $x \in A$ ant let (U, h) be an element of the atlas for M such that $x \in U$. Then since A is open so is $h(U \cap A \subset \mathbb{R}^n$. Hence there is an open ball $B \subset U \cap A$ centered at $h(x)$. let $V := h^{-1}(B)$ and h' be the restriction of h to V. By Remark 4.1.4 pairs (V, h') obtained this way from various points $x \in A$ form an atlas compatible with the differentiable structure on M. □

Smooth manifolds constitute a category, whose morphisms are appropriately called smooth (or differentiable) maps. An important class of smooth maps of a fixed manifold is the class of its maps to $\mathbb{R}$, or smooth functions. We will see that smooth functions form an $\mathbb{R}$-algebra from which the manifold can be entirely reconstructed.

A real-valued function $f : M \to \mathbb{R}$ on a smooth manifold M is called *smooth* (or *differentaible*) if on each chart (U, h) the composition $f \circ h^{-1}$ is a differentiable function from $\mathbb{R}^n$ to $\mathbb{R}$. Using the compatibility condition, it is easy to verify that it suffices to check differentiability for any set of charts covering M (rather than for all charts of its maximal atlas). maybe we sho

The set of all smooth functions on M will be denoted by $C^\infty(M)$ (or $C^n(M)$, $n \in \mathbb{N}$, depending on the differentiability class under consideration).

One of the remarkable mathematical discoveries of the mid-twentieth century was the realization that a topological manifold can have more than one differentiable structure: even the sphere (e.g. in dimension 7) can have several different smooth structures. Further, certain topological manifolds have no smooth structure compatible with their topology. These delicate questions will not be discussed in this course.

4.1.2. First examples.

EXAMPLE 4.1.7. $\mathbb{R}^n$ is a smooth manifold with an atlas consisting of a single chart: the identity of $\mathbb{R}^n$.

Any open subset of $\mathbb{R}^n$ is also an n-dimensional differentiable manifold by Proposition 4.1.6. However, it may not be diffeomorphic to $\mathbb{R}^n$ and hence in general would not possess an atlas with single chart.

EXAMPLE 4.1.8. An interesting specific example of this kind is obtained by viewing the linear space of $n \times n$ matrices as $\mathbb{R}^{n^2}$. The condition $\det A \neq 0$ then defines an open set, hence a manifold (of dimension n^2), which is familiar as the *general linear group* $GL(n, \mathbb{R})$ of invertible $n \times n$ matrices.

EXERCISE 4.1.2. Construct an explicit diffeomorphism between $\mathbb{R}^n$ and the open unit ball B^n.

EXERCISE 4.1.3. Prove that any convex open set in $\mathbb{R}^n$ is diffeomorphic to $\mathbb{R}^n$.

se is better turned
position; at least it
s an extensive hint

EXAMPLE 4.1.9. The standard sphere $\mathbb{S}^2 \subset \mathbb{R}^3$ is a differentiable manifold. As charts modeled on the open ball one can take six appropriately chosen parallel projections of hemispheres to the coordinate planes. More economically, one gets a cover by two charts modeled on $\mathbb{R}^2$ by the two stereographic projections of the sphere from its north and south poles. As a forward reference we notice that if $\mathbb{R}^2$ is identified with $\mathbb{C}$ the latter method also provides $\mathbb{S}^2$ with the structure of one-demansional complex manifold (see Section 4.9.1).

EXAMPLE 4.1.10. The embedded torus

$$\mathbb{T}^2 = \left\{(x, y, z) \in \mathbb{R}^3 : \left(\sqrt{x^2 + y^2} - 2\right)^2 + z^2 = 1\right\}$$

can be covered by overlapping pieces of parametrized surfaces

$$W \ni (u, v) \mapsto \big(x(u, v), y(u.v), z(u, v)\big) \in U \subset \mathbb{T}^2$$

whose inverses $U \to W$ (see the figure) constitute an atlas of $\mathbb{T}^2$, so it has the structure of a two-dimensional smooth manifold.

FIGURE ??? Chart on the embedded torus

EXERCISE 4.1.4. Using the square with identified opposite sides as the model of the torus, construct a smooth atlas of the torus with four charts homeomorphic to the open disk.

EXERCISE 4.1.5. Construct a smooth atlas of the projective space $\mathbb{R}P(3)$ with as few charts as possible.

EXAMPLE 4.1.11. The surface of a regular tetrahedron can be endowed with the structure of a two-dimensional smooth manifold by embedding it into 3-space, projecting it from its center of gravity G onto a 2-sphere of large radius centered at G, and pulling back the charts of the sphere to the surface.

Intuitively, there is something unnatural about this smooth structure, because the embedded tetrahedron has "corners", which are not "smooth" in the everyday sense. We will see below that a rigorous definition corresponds to this intuitive feeling: the embedded tetrahedron is not a "submanifold" of $\mathbb{R}^3$ (see Definition 4.2.1).

4.1.3. Manifolds defined by equations. Joint level sets of smooth functions into $\mathbb{R}$ or $\mathbb{R}^m$ corresponding to regular values are an interesting general class of manifolds. This is the most classical source of examples of manifolds.

Charts are provided by the implicit function theorem. Due to importance of this method we will give a detailed exposition here.

THEOREM 4.1.12 (Implicit Function Theorem). *Let $O \subset \mathbb{R}^n \times \mathbb{R}^m$ be open and $f\colon O \to \mathbb{R}^n$ a C^r map. If there is a point $(a,b) \in O$ such that $f(a,b) = 0$ and $D_1 f(a,b)$ is invertible then there are open neighborhoods $U \subset O$ of (a,b), $V \subset \mathbb{R}^m$ of b such that for every $y \in V$ there exists a unique $x =: g(y) \in \mathbb{R}^n$ with $(x,y) \in U$ and $f(x,y) = 0$. Furthermore g is C^r and $Dg(b) = -(D_1 f(a,b))^{-1} D_2 f(a,b)$.*

Proof of this theorem can be found in ??.

Examples are the sphere in $\mathbb{R}^n$ (which is the level set of one function, e.g. $F(x,y,z) = x^2 + y^2 + z^2$, for which 1 is a regular value) and the *special linear group* $SL(n,\mathbb{R})$ of $n \times n$ matrices with unit determinant. Viewing the space of $n \times n$ matrices as $\mathbb{R}^{n^2}$, we obtain $SL(n,\mathbb{R})$ as the manifold defined by the equation $\det A = 1$. One can check that 1 is a regular value for the determinant. Thus this is a manifold defined by one equation.

4.2. Principal constructions

Now we will look at how the notion of smooth manifold behaves with respect to the basic constructions. This will provide as with two principal methods of constructing smooth manifolds other than directly describing an atlas: embeddings as submanifolds, and projections into factor-spaces.

4.2.1. Submanifolds. In the case of a topological or a metric space, any subset automatically acquires the corresponding structure (induced topology or metric). For smooth manifolds, the situation is more delicate: arbitrary subsets of a smooth manifold do not necessarily inherit a differentiable structure from the ambient manifold. The following definition provides a natural generalization of the notion described in the previous subsection.

DEFINITION 4.2.1. A *submanifold* V of M (of dimension $k \leq n$) is a differentiable manifold that is a subset of M such that the maximal atlas for M contains charts $\{(U,h)\}$ for which the restrictions $h_{\restriction U \cap V}$ map $U \cap V$ to $\mathbb{R}^k \times \{0\} \subset \mathbb{R}^n$ define charts for V compatible with the differentiable structure of V.

EXAMPLE 4.2.2. An open subset of a differentiable manifold M with the induced atlas as described in Proposition 4.1.6 is a submanifold of dimension n.

EXAMPLE 4.2.3. Let C be simple closed polygonal curve in $\mathbb{R}^2$ and let $h : C \to \mathbb{S}^1$ be a homeomorphism; then C acquires a smooth structure (via the atlas $\mathcal{A}_h$, the pullback by h of the standard atlas of $\mathbb{S}^1$). The curve C with this smooth structure is *not* a submanifold of the smooth manifold $\mathbb{R}^2$ (because of the "corners"). The same can be said of the tetrahedron embedded in 3-space, see Exercise 4.1.11.

EXERCISE 4.2.1. Prove that the n-dimensional torus in $\mathbb{R}^{2n}$:

$$x_{2k-1}^2 + x_{2k}^2 = \frac{1}{n}, \quad k = 1, \dots, n$$

is a smooth submanifold of the $(2n-1)$-dimensional sphere

$$\sum_{i=1}^{2n} x_i^2 = 1.$$

EXERCISE 4.2.2. Prove that the upper half of the cone

$$x^2 + y^2 = z^2, \;\; z \geq 0$$

is not a submanifold of $\mathbb{R}^3$, while the punctured one

$$x^2 + y^2 = z^2, \;\; z > 0$$

is a submanifold of $\mathbb{R}^3$.

Conversely, every smooth n-manifold can be viewed as a submanifold of R^N for a large enough N (see Theorem 4.5.1 and **??**).

4.2.2. Direct products. The Cartesian product of two smooth manifolds M and N of dimensions m and n automatically acquires the structure of an $(n+m)$–dimensional manifold in the following (natural) way. In the topological space $M \times N$, consider the atlas consisting of the products $U_i \times V_j$ of all pairs of charts of M and N with the natural local coordinates

$$l_{ij} := h_i \times k_j : U_i \times V_j \to \mathbb{R}^{m+n}.$$

It is easy to see that these charts are compatible and constitute an atlas of $M \times N$.

EXERCISE 4.2.3. Show that the smooth structure obtained on the torus $\mathbb{T}^2 = \mathbb{S}^1 \times \mathbb{S}^1$ in the above way coincides with that induced from the standard embedding of the torus in 3-space.

continuous actions
phisms (discrete);
s with continuous
(implicit function)

4.2.3. Quotient spaces. Identification spaces can also be smooth manifolds, for example, the unit circle viewed as $\mathbb{R}/\mathbb{Z}$, the torus as $\mathbb{R}^n/\mathbb{Z}^n$, or compact factors of the hyperbolic plane **??**.

Note that, conversely, given a covering map of a smooth manifold, its smooth structure always lifts to a smooth structure of the covering space.

EXERCISE 4.2.4. Prove that the following three smooth structures on the torus $\mathbb{T}^2$ are equivalent, i.e. the torus provided with any of these structure is diffeomorphic to the one provided with another:

- $\mathbb{T}^2 = \mathbb{S}^1 \times \mathbb{S}^1$ with the product structure;
- $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$ with the factor-structure;
- The embedded torus of revolution in $\mathbb{R}^3$
$$\mathbb{T}^2 = \left\{(x, y, z) \in \mathbb{R}^3 : \left(\sqrt{x^2 + y^2} - 2\right)^2 + z^2 = 1\right\}$$
with the submanifold structure.

4.3. Orientability and degree

4.3.1. Orientation and orientability.

4.3.2. Easy part of Sard theorem.

4.3.3. Degree for maps of compact orientable manifolds.

4.3.4. Calculation of $\pi_n(\mathbb{S}^n)$.

4.4. Paracompactness and partition of unity

An important result for analysis on manifolds is the fact that (using our assumption of second countability, that is, that there is a countable base for the topology) every smooth manifold admits a partition of unity (used below, in particular, to define the volume element of a manifold), which is defined as follows.

DEFINITION 4.4.1. A *partition of unity* subordinate to a cover $\{U_i\}$ of a smooth manifold M is a collection of continuous real-valued functions $\varphi_i : M \to [0, 1]$ such that

• the collection of functions φ_i is *locally finite*, i.e., any point $x \in M$ has a neighborhood V which intersects only a finite number of sets $\operatorname{supp}(\varphi_i)$ (recall that the support of a function is the closure of the set of points at which it takes nonzero values);
• $\sum_i \varphi_i(x) = 1$ for any $x \in M$;
• $\operatorname{supp}(\varphi) \subset U_i$ for all i.

PROPOSITION 4.4.2. *For any locally finite cover of a smooth manifold M, there exists a partition of unity subordinate to this cover.*

PROOF. Define the functions $g_i : M \to [0, 2^{-i}]$ by setting

if we need
partition of uni
a

$$g_i := \min\{d(x, M \setminus U_i,\, 2^{-i}\},$$

where $d(\cdot\,,\,\cdot)$ denotes the distance between a point and a set and $\{U_i\}$ is the given cover of M. Then we have $g_i(x) > 0$ for $x \in U_i$ and $g_i(x) = 0$ for $x \notin U_i$. Further define

$$G(x) := \lim_{N\to\infty} G_N(x) = \lim_{N\to\infty} \sum_{i=1}^{N} f_i(x).$$

Since $\{U_i\}$ is a cover, it follows that $G(x) > 0$ for all $x \in M$.

Now put

$$f_i(x) := \max\Big\{g_i(x) - \frac{1}{3}G(x), 0\Big\}.$$

It is then easy to see that $\operatorname{supp}(f_i) \subset U_i$, and, since the cover $\{U_i\}$ is locally finite, so is the system of functions $\{f_i\}$.

Now let us show that

$$F(x) := \sum_{i=1}^{\infty} f_i(x) > 0 \quad \text{for all} x \in M,$$

i.e., for any $x \in M$ there is an i for which $f_i(x) > 0$. We do know that $g_j(x > 0)$ for some j and $g_n(x) < 2^{-n}$, hence $\sup_{j\in\mathbb{N}} g_j(x) = g_{i_0}(x)$ for a certain i_0 such that $g_{i_0} > 0$. The definition of the function $G(x)$ implies

$$G(x) = \sum_{j=0}^{\infty} 2^{-j} g_j(x) \le \sum_{j=0}^{\infty} 2^{-j} g_{i_0}(x) = 2g_{i_0}.$$

Therefore

$$f_{i_0}(x) \ge g_{i_0}(x) - \frac{2g_{i_0}(x)}{3} = \frac{g_{i_0}(x)}{3} > 0.$$

Now we can define the required partition of unity by setting

$$\varphi(x) := f_i(x)/F(x).$$

The proof of the facts that the φ_i are continuous, form a locally finite family, and add up to 1 at any point $x \in M$ is a straightforward verification that we leave to the reader. $\square$

A topological space X is called *paracompact* if a locally finite open cover can be inscribed in in any open cover of X.

PROPOSITION 4.4.3. *Any smooth manifold M is paracompact.*

PROOF. Let $\{U_i\}$ be an open cover of M, which we assume countable without loss of generality. Then the interiors of the supports of the functions φ_i obtained by the construction (which does not use the local finiteness of the covering $\{U_i\}$) in the proof of the previous proposition will form a locally finite open cover of M subordinated to $\{U_i\}$. $\square$

COROLLARY 4.4.4. *Any smooth manifold possesses a locally finite cover with a partition of unity subordinate to it.*

4.5. Embedding into Euclidean space

In this section we will prove that any compact differentiable manifold is diffeomorphic to a submanifold of a Euclidean space of a sufficiently high dimension.

THEOREM 4.5.1. *Any smooth compact manifold M^n can be smoothly embedded in Euclidean space $\mathbb{R}^N$ for sufficiently large N .*

Why only compa
is true for

PROOF. Since the manifold M^n is compact, it possesses a finite family of charts $f_i : U_i \to \mathbb{R}^n$, $i = 1, \dots, k$, such that

(1) the sets $f_i(U_i)$ are open balls of radius 2 centered at the origin of $\mathbb{R}^n$;

(2) the inverse images (denoted V_i) by f_i of the unit balls centered at the origin of $\mathbb{R}^n$ cover M^n.

We will now construct a smooth "cut off" function $\lambda : \mathbb{R}^n \to \mathbb{R}$ such that

$$\lambda(x) = \begin{cases} 1 & \text{for} \quad \|y\| \le 1, \\ 0 & \text{for} \quad \|y\| \ge 2, \end{cases}$$

and $0 < \lambda(y) < 1$ for $1 < \|y\| < 2$. To do this, we first consider the function

$$\alpha(x) := \begin{cases} 0 & \text{for} \quad \|x\| \le 0, \\ e^{-1/x} & \text{for} \quad \|x\| > 0, \end{cases}$$

and then put $\beta(t) := \alpha(x-1)\alpha(2-x)$; the function β is positive on the open interval $(1, 2)$. Finally, we define

$$\gamma(\tau) := \Big(\int_\tau^2 \beta(t)dt \Big) \Big/ \Big(\int_1^2 \beta(t)dt \Big)$$

and put $\lambda(y) := \gamma(\|y\|)$. This function obviously satisfies the conditions listed above.

We set $\lambda_i(x) := \lambda(f_i(x))$ (see the figure).

FIGURE ??? The cut off function λ_i

Now let us consider the map $h : M^n \to \mathbb{R}^{(n+1)k}$ given by the formula

$$x \mapsto \big(\lambda_1(x),\, \lambda_1(x)f_1(x), \dots, \lambda_k(x), \lambda_k(x)f_k(x)\big).$$

The map h is one-to-one. Indeed, let $x_1, x_2 \in M^n$. Then x_1 belongs to V_i for some i, and two cases are possible: $x_2 \in V_i$ and $x_2 \notin V_i$. In the first case, we have $\lambda_i(x_1) = \lambda_i(x_2) = 1$, and therefore the relation

$$\lambda_i(x_1) f_i(x_1) = \lambda_i(x_2) f_i(x_2)$$

is equivalent to $f_i(x_1) = f_i(x_2)$ and so $x_1 = x_2$. In the second case (when $x_2 \notin V_i$), we have $\lambda_i(x_1) = 1$ while $\lambda_i(x_2) < 1$, and so $h(x_1) \neq h(x_2)$.

Now the restriction of the map $x \mapsto \lambda_i(x) f_i(x)$ to U_i is an immersion (i.e., at any point its Jacobian is of rank n), because the inclusion $x \in U_i$ implies $\lambda_i(x) = 1$), while the map $x \mapsto f_i(x)$ is a local diffeomorphism. Hence the map h is also an immersion.

But we know (see ??) that any one-to-one map of a compact space into a Hausdorff space (in our case $h : M^n \mathbb{R}^{(n+1)k}$) is a homeomorphism onto its image. Thus h is a smooth embedding into $\mathbb{R}^{(n+1)k}$. □

4.6. Derivatives and the tangent bundle

4.6.1. Derivations as classes of curves. Recall that the derivative of a function $f : \mathbb{R}^n \to \mathbb{R}$ in the direction of a vector $v = (v_1, \dots, v_n)$ is defined in calculus courses as

$$D_v(f) := v_1 \frac{\partial f}{\partial x_1} + \cdots + v_n \frac{\partial f}{\partial x_n}.$$

Derivations form a linear space of dimension n whose canonical basis is constituted by the partial derivatives

$$\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n}.$$

In order to give a similar definition of the derivative of a function on a smooth manifold, we must, first of all, define what we mean by the direction along which we differentiate. We will do this by defining tangent vectors as equivalence classes of curves. The underlying intuitive consideration is that curves passing through a point are viewed as trajectories, two curves being regarded as equivalent if the "velocity of motion" at the chosen point is the same.

DEFINITION 4.6.1. Let M be a C^∞ manifold and $p \in M$. Consider curves $c\colon (a, b) \to M$, where $a < 0 < b$, $c(0) = p$ such that $h \circ c$ is differentiable at 0 for one (hence any) chart (U, h) with $p \in U$. Each such curve c passing through the point p assigns to each function $f \in C^\infty(M)$ the real number

$$D_{c,p}(f) := \frac{d}{dt}\big(f(c(t))\big|_{t=0},$$

the *derivative of f at p along c*. Two curves c' and c'' are called *equivalent* if in some chart (U, h) (and hence, by compatibility, in all charts) containing p, we have

$$\frac{d}{dt}\big(h(c'(t))\big|_{t=0} = \frac{d}{dt}\big(h(c''(t))\big|_{t=0}.$$

An equivalence class of curves at the point p is called a *tangent vector* to M at p and denoted by $v = v(c)$, where c is any curve in the equivalence class. The *derivative of f in the direction of the vector v* can now be (correctly!) defined by the formula

$$D_{v,p}(f) := \frac{d}{dt}\big(f(c(t))\big|_{t=0}, \quad \text{for any } c \in v.$$

The space of all the derivations at p (i.e., equivalence classes of curves at p) obtained in this way, has a linear space structure (since each derivation is a real-valued function) which turns out to have dimension n. It is called the *tangent space* at p of M and denoted T_pM.

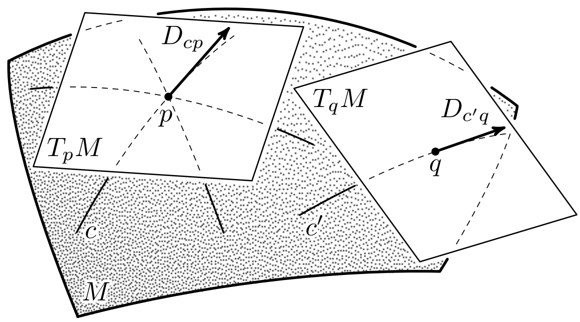

FIGURE 4.6.1. Tangent spaces to a manifold

Given a specific chart (U, h), we define the standard basis

$$\frac{\partial}{\partial x_1}\Big|_p, \dots, \frac{\partial}{\partial x_n}\Big|_p$$

of T_pM by taking the canonical basis $\{e_1, \dots, e_n\}$ of $\mathbb{R}^n$ and setting

$$\frac{\partial}{\partial x_i}\Big|_p(f) := \frac{d}{dt}\big(f(c_i(t))\big)\Big|_{t=0}, \quad \text{where } c_i(t) = h^{-1}(h(p) + te_i)$$

for all $i = 1, \dots, n$.

4.6.2. Derivations as linear operators. Another intrinsic way of defining derivatives, more algebraic than the geometric approach described in the previous subsection, is to define them by means of linear operators satisfying the Leibnitz rule.

DEFINITION 4.6.2. Let p be a point of a smooth manifold M. A *derivation* of $C^\infty(M)$ at the point p is a linear functional $D : C^\infty(M) \to \mathbb{R}$ satisfying the Leibnitz rule, i.e.,

$$D(f \cdot g) = Df \cdot g(p) + f(p) \cdot Dg.$$

The derivations at p (in this sense) obviously constitute a linear space. If we choose a fixed chart (U, h) with coordinates $(x_1, \dots, x_n)$ containing p, then we can determine a basis $(\partial_1, \dots, \partial_n)$ of this space by setting

$$D_1(f) := \frac{\partial}{\partial x_1}(h \circ f)\Big|_{f(p)}, \dots, D_n(f) := \frac{\partial}{\partial x_2}(h \circ f)\Big|_{f(p)};$$

here $\partial/\partial x_i$ denotes the usual partial derivative in the target space $\mathbb{R}^n$ of our chart h.

EXERCISE 4.6.1. Prove that the linear space of derivations can be identified with the tangent space $T_p(M)$ defined in the previous subsection, so that the derivations defined above are nothing but tangent vectors and the basis $\{D_i\}$ can be identified with the basis $\{(\partial/\partial x_i)|_p\}$.

REMARK 4.6.3. Note that the definition of derivation given in this subsection yields a purely algebraic approach to the differential calculus on smooth manifolds: none of the classical tools of analysis (e.g. limits, continuity via the $\varepsilon-\delta$ language, infinite series, etc.) are involved.

4.6.3. The tangent bundle. We define the *tangent bundle* of M to be the disjoint union

$$TM := \bigcup_{p\in m} T_pM$$

of the tangent spaces with the *canonical projection* $\pi\colon TM \to M$ given by $\pi(T_pM) = \{p\}$. Any chart (U,h) of M then induces a chart

$$\Big(U \times \bigcup_{p\in U} T_pU, H\Big), \quad \text{where } H(p,v) := (h(p),(v^1,\dots,v^n)) \in \mathbb{R}^n \times \mathbb{R}^n;$$

here the v^i are the coefficients of $v \in T_pM$ with respect to the basis

$$\left\{\frac{\partial}{\partial x^1}\Big|_p, \dots, \frac{\partial}{\partial x^n}\Big|_p\right\}$$

of T_pM. In this way TM is a differentiable manifold (of dimension 2) itself.

A *vector field* is a map $X\colon M \to TM$ such that $\pi \circ X = \mathrm{Id}_M$, that is, X assigns to each p a tangent vector at p. We denote by $\Gamma(M)$ the space of *smooth* vector fields on M, i.e., vector fields defined by a smooth map of the manifold M to the manifold TM. Thus smooth vector fields determine operators (that we will sometimes denote by D_X) on $C^\infty(M)$ by acting on functions via derivations, i.e., $D_X(f) := X(p)(f)$.

We shall see later that $\pounds_v w := [v,w] := vw - wv$ also acts on functions by derivations, that is, as a vector field, and we call $[v,w]$ the *Lie bracket* of v and w and $\pounds_v$ the *Lie derivative* .

4.7. Smooth maps and the tangent bundle

As we already noted, smooth manifolds, like any other self-respecting mathematical objects, form a category: we defined their morphisms (called smooth maps) and their "isomorphisms" (called diffeomorphisms) at the beginning of the present chapter. We now return to these notions and look at them from the perspective of tangent bundles.

4.7.1. Main definitions. We now define the morphisms of the differentiable structure.

DEFINITION 4.7.1. Let M and N be differentiable manifolds. Recall that a map $f\colon M \to N$ is said to be *smooth* if for any charts (U, h) of M and (V, g) of N the map $g \circ f \circ h^{-1}$ is differentiable on $h(U \cap f^{-1}(V))$.

A smooth map f acts on derivations by sending curves $c\colon (a, b) \to M$ to $f \circ c\colon (a, b) \to N$. Differentiability means that curves inducing the same derivation have images inducing the same derivation. Thus we define the differential of f to be the map

$$Df\colon TM = \bigcup_{p \in M} T_pM \to TN = \bigcup_{q \in N} T_qN$$

that takes each vector $v \in T_pM$ determined by a curve c to the vector $w \in T_{f(p)}$ given by the curve $f \circ c$. It is easy to deduce from the definition of equivalence of curves (see ??) that the definition of w does not depend on the choice of curve $c \in v$. The restriction of Df to T_pM (which takes T_pM to $T_{f(p)}N$) is denoted by $Df\big|_p$.

A *diffeomorphism* is a differentiable map with differentiable inverse. Two manifolds M, N are said to be *diffeomorphic* or *diffeomorphically equivalent* if and only if there is a diffeomorphism $M \to N$. An *embedding*of a manifold M in a manifold N is a diffeomorphism $f\colon M \to V$ of M onto a submanifold V of N. We often abuse terminology and refer to an embedding of an open subset of M into N as a (*local*) *diffeomorphism* as well. An *immersion* of a manifold M into a manifold N is a differentiable map $f\colon M \to V$ onto a subset of N whose differential is injective everywhere.

4.7.2. Examples. Smooth maps must be compatible, in a sense, with the differentiable structure of the source and target manifolds. As we shall see, not all naturally defined maps (e.g. some projections) have this property.

EXAMPLE 4.7.2. The orthogonal projection on the (x, y)-plane of the standard unit sphere $x^2 + y^2 + z^2 = 1$ is not a smooth map.

Further, even injectively immersed manifolds may fail to be smooth submanifolds.

EXAMPLE 4.7.3. Choose a point on the standard embedding of the torus $\mathbb{T}^2$ and consider a curve passing through that point and winding around $\mathbb{T}^2$ with irrational slope (forming the same irrational angle at all its intersections with the parallels of the torus). In that way, we obtain a (dense) embedding of $\mathbb{R}$ into $\mathbb{T}^2$, which is a smooth map *locally*, but is *not* a smooth map of $\mathbb{R}$ to $\mathbb{T}^2$.

Clearly, diffeomorphic manifolds are homeomorphic. The converse is, however, not true. As we mentioned above, there are "exotic" spheres and other manifolds whose smooth structure is not diffeomorphic to the usual smooth structure.

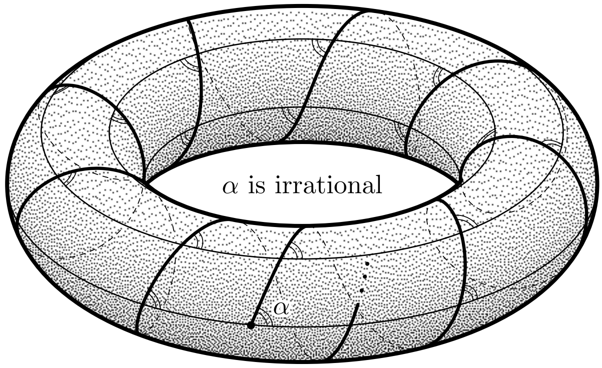

FIGURE 4.7.1. Dense embedded trajectory on the torus

EXAMPLE 4.7.4. In the space $\mathbb{R}^9$ with coordinates $(x_1, \dots, x_9)$, consider the cone C given by

$$x_1^7 + 3x_7^4 x_2^3 + x_5^6 x_6 = 0$$

and take the intersection of C with the standard unit 8-sphere $\mathbb{S}^8 \subset \mathbb{R}^9$. The intersection $\Sigma := C \cap \mathbb{S}^8$ is clearly homeomorphic to the 7-sphere. It turns out that Σ with the smooth structure induced on from $\mathbb{R}^9$ is not diffeomorphic to $\mathbb{S}^7$ with the standard smooth structure. (The proof of this fact lies beyond the scope of the present book.)

4.8. Manifolds with boundary

The notion of real smooth manifold with boundary is a generalization of the notion of real smooth manifold obtained by adding the half-space

$$\mathbb{R}^n_+ := \{(x_1, \dots, x_n \in \mathbb{R}^n \mid x_n \geq 0\}$$

to $\mathbb{R}^n$ as the possible target space of the charts (U_i, h_i); we must also appropriately modify the compatibility condition: we now require that, whenever $U_r \cap U_s \neq \varnothing$, there must exist two mutually inverse diffeomorphisms $\varphi_{r,s}$ and $\varphi_{s,r}$ of *open* sets in $\mathbb{R}^n$ whose restrictions are $h_r \circ h_s^{-1}$ and $h_s \circ h_r^{-1}$. (The necessity of such a version of the compatibility condition is in that smooth maps are defined only on open subsets of $\mathbb{R}^n$, whereas an open set in $\mathbb{R}^n_+$, e.g. $h_i(U_i)$, may be non open in $\mathbb{R}^n$.)

If M is a smooth manifold with boundary, then it has two types of points: the *interior* points (those contained in only in those charts (U_i, h_i) for which $U_i \subset M$ is open) and the *boundary* points (those not contained in any such charts). It seems obvious that the *boundary* ∂M of a manifold with boundary (i.e., the set of its boundary points) coincides with the set

$$\bigcup_j h_j^{-1}((x_1, \dots, x_{n-1}, 0)),$$

where the intersection is taken over only those h_j whose target space is $\mathbb{R}^n_+$. However, this fact is rather nontrivial, and we state it as a lemma.

LEMMA 4.8.1. *The two definitions of boundary point of a manifold with boundary coincide.*

PROOF. We need to prove that any point contained in an *open* chart U_i cannot be mapped by h_i to a boundary point $(x_1, \dots, x_{n-1}, 0) \in \mathbb{R}^n_+$. This can be done by using the inverse function theorem. We omit the details. $\square$

Sometimes, in order to stress that some M is an ordinary manifold (not a manifold with boundary), we will say that M is a "manifold without boundary". It may happen that the set of boundary points of a manifold with boundary M is empty. In that case, all the charts of its maximal atlas targeted to $\mathbb{R}^n_+$ are in fact redundant; deleting them, we obtain a smooth manifold without boundary.

PROPOSITION 4.8.2. *The set of boundary points ∂M of a manifold with boundary has the natural structure of a smooth $(n-1)$-dimensional manifold (without boundary).*

PROOF. An atlas for ∂M is obtained by taking the restrictions of the charts h_i to the sets $h_i^{-1}(h_i(U_i) \cap \mathbb{R}^n_+)$. $\square$

4.9. Complex manifolds

4.9.1. Main definitions and examples. Complex manifolds are defined quite similarly to real smooth manifolds by considering charts with values in $\mathbb{C}^n$ instead of $\mathbb{R}^n$ and requiring the coordinate changes between charts to be holomorphic. Since holomorphic maps are much more rigid that differentiable maps, the resulting theory differs from the one above in several aspects. For example the one–dimensional complex manifolds (Riemann surfaces) is a much richer subject than one- and even two-dimensional differentiable manifolds.

Complex manifolds form a category, the natural notion of morphism $\varphi : M \to N$ being defined similarily to that of smooth map for their real counterparts, except that the maps $k \circ \varphi h^{-1}$ (where h and k are charts in M and N) must now be holomorphic rather than differentiable.

In this course, we do not go deeply into the theory of complex manifolds, limiting our study to some illustrative examples.

EXAMPLE 4.9.1. The *Riemann sphere*, $\mathbb{C} \cup \{\infty\}$, which is homeomorphic to S^2, becomes a one–dimensional complex manifold by considering an atlas of two charts $(\mathbb{C}, \mathrm{Id})$ and $(\mathbb{C} \cup \{\infty\} \setminus \{0\}, I)$, where

$$I(z) = \begin{cases} 1/z & \text{if } z \in \mathbb{C} \\ 0 & \text{if } z = \infty \end{cases}.$$

EXERCISE 4.9.1. Identify $\mathbb{R}^2$ with $\mathbb{C}$ and define the torus $\mathbb{T}^2$ as the quotient space $\mathbb{C}/\mathbb{Z}^2$.

EXERCISE 4.9.2. Describe a complex atlas for the complex projective space $\mathbb{C}P^n$.

EXERCISE 4.9.3. Describe a complex atlas for the group $U(n)$ of unitary matrices

4.9.2. Riemann surfaces. An attractive showcase of examples of complex manifolds comes from *complex algebraic curves* (or *Riemann surfaces*, as they are also called), which are defined as zero sets of complex polynomials of two variables in the space $\mathbb{C}^2$.

More precisely, consider the algebraic equation

$$p(z,w) := a_0(z)w^n + a_1(z)w^{n-1} + \cdots + a_n(z) = 0, \quad a_0(z) \neq 0, \tag{4.9.1}$$

where the $a_i(z)$ are polynomials in the complex variable $z \in \mathbb{C}$ with complex coefficients and $w = w(z)$ is an unknown complex-valued function.

Already in the simplest cases (e.g. for $w^2 - z = 0$), this equation does not have a univalent analytic solution $w : \mathbb{C} \to \mathbb{C}$ defined for all $z \in \mathbb{C}$. However, as Riemann noticed, such a solution exists provided we replace the domain of definition of the solution by an appropriately chosen surface that we will now define.

To do this, it will be convenient to replace $\mathbb{C}$ by its natural compactification $\bar{\mathbb{C}} := \mathbb{C} \cup \{\infty\}$, the *Riemann sphere*, which is of course homeomorphic to the ordinary sphere $\mathbb{S}^2$). We now regard equation (4.5.1.) as given on $\bar{\mathbb{C}} \times \bar{\mathbb{C}}$ and define the corresponding *Riemann surface* S_p as the set of zeros of this equation, i.e., as

$$S_p := \big\{(z,w) \in \bar{\mathbb{C}} \times \bar{\mathbb{C}} \,|\, p(z,w) = 0\big\}.$$

Now the projection (given by the assignment $(z,w) \mapsto w$) of S_p on the second factor of the product $\bar{\mathbb{C}} \times \bar{\mathbb{C}}$ is by definition univalent, so that on the Riemann surface S_p equation (4.5.1.) defines a single-valued function $w = w(z)$.

It is of course difficult to visualize Riemann surfaces, which are two-dimensional objects embedded in a four-dimensional manifold homeomorphic to $\mathbb{S}^2 \times \mathbb{S}^2$, but we will see that there is an effective geometric construction that, given $p(z,w)$, specifies the topological structure of S_p.

We will now consider several examples of this construction.

EXAMPLE 4.9.2. Consider the equation

$$p(z) := w^2 - z = 0.$$

Obviously, there are two values of w that satisfy this equation for a fixed (nonzero) value of $z = re^{\varphi}$, namely $w_1 = +\sqrt{r}\, e^{i\varphi/2}$ and $w_2 = -\sqrt{r}\, e^{i\varphi/2}$. These determine the two "sheets" of the solution; when we go around the origin of the z-plane, we "jump" from one sheet to the other. Let us cut the z-plane along the real axis, or more precisely cut the Riemann sphere $\bar{\mathbb{C}}$ along the arc arc of the great circle joining the points 0 and ∞. Take another copy of $\bar{\mathbb{C}}$ (which will be the second sheet of our Riemann surface), make the same cut joining 0 and ∞, and identify the "shores" of the cuts (see the figure below).

Thus we see that the Riemann surface of the equation under consideration is the sphere.

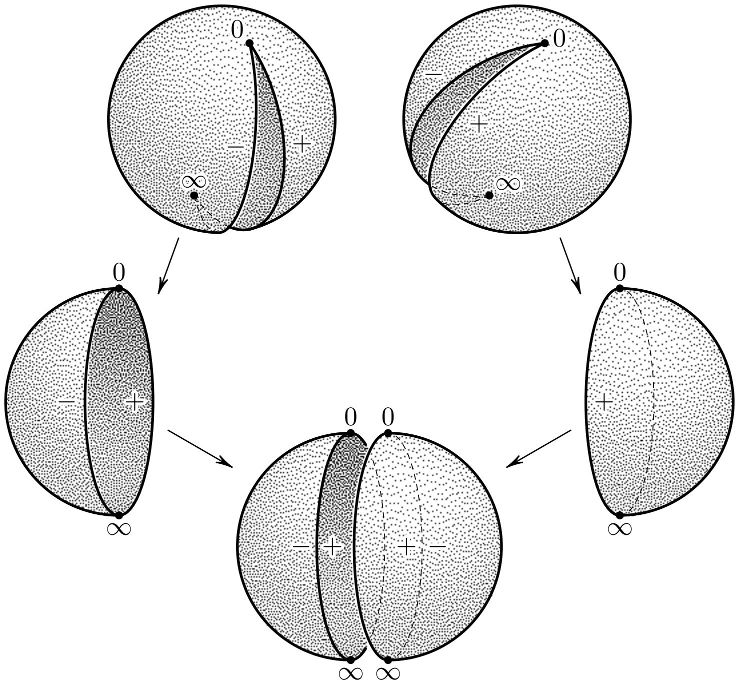

FIGURE 4.9.1. The Riemann surface of a polynomial linear in z

EXERCISE 4.9.4. Show that the Riemann surface of the quadratic equation

$$w^2 - (z - a_1)(z - a_2) = 0,$$

where a_1 and a_2 are distinct complex numbers, is the sphere $\S^2_b$.

EXAMPLE 4.9.3. Consider the cubic equation

$$q(z, w) := w^2 - (z - a_1)(z - a_2)(z - a_3) = 0,$$

where a_1, a_2, a_3 are distinct complex numbers. This function also has two sheets, but the passage from one sheet to the other is more complicated than in the previous example: if we circle around one of the points a_1, a_2, a_3, or ∞, we pass from one sheet to the other, if we circle around any two of them, we stay on the same sheet, if we circle around three, we switch sheets again. To obstruct these switches, we perform cuts along the arcs a_1a_2 and $a_3\infty$ on two copies of the Riemann sphere and glue the two copies along the shores of the cuts. The construction is shown on the figure.

The result will clearly be homeomorphic to the torus.

EXERCISE 4.9.5. Find a polynomial whose zero set is a complex curve homeomorphic to the sphere with two handles.

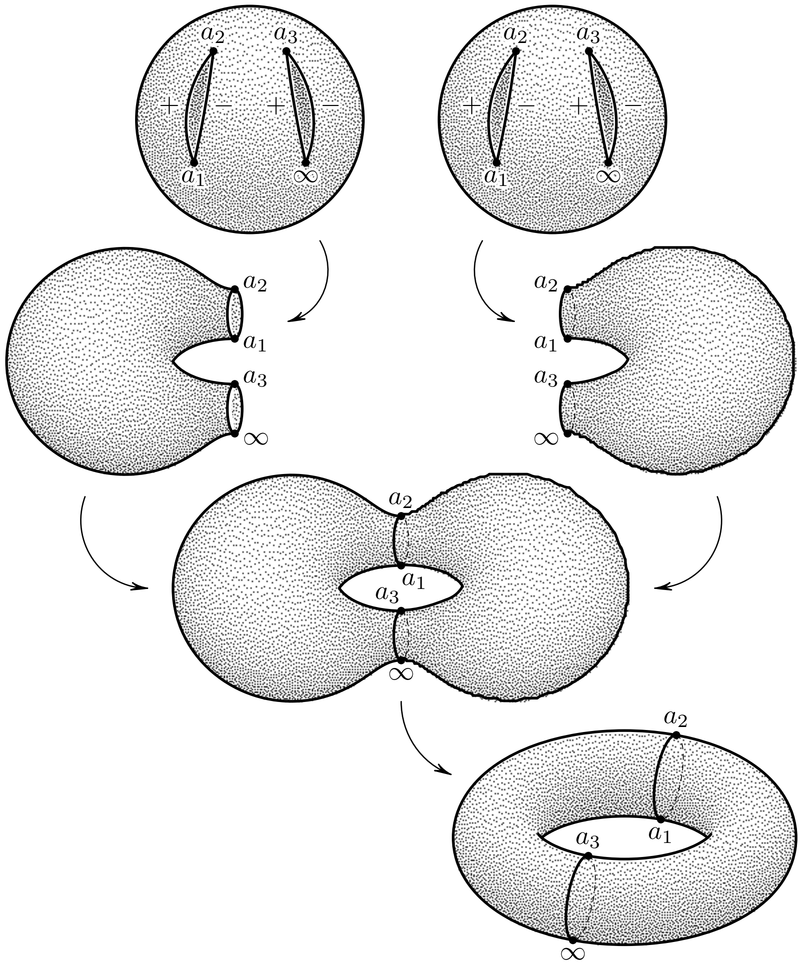

FIGURE 4.9.2. The Riemann surface of a polynomial cubic in z

4.10. Lie groups: first examples

DEFINITION 4.10.1. An *n-dimensional Lie group* is an n-dimensional smooth manifold G with a group operation such that the product map $G \times G \to G\colon (x, y) \mapsto xy$ and the inverse map $G \to G\colon x \mapsto x^{-1}$ are differentiable.

Lie groups G and H are *isomorphic* is there exists a group isomorphism $i\colon G \to H$ which is at the same time a diffeomorphism between smooth manifolds.

A *Lie subgroup* of a Lie group G is a smooth submanifold H of G which is also a subgroup. [2]

Lie groups form one of the most important and interesting classes of smooth manifolds. Here we discuss few examples of the classical Lie groups and mention

[2] In fact, any closed subgroup of a Lie group is a Lie subgroup. This is one of the fundamental results of the Lie group theory which is used quite often. Its proof is far from elementary.

some of their properties. More systematic study of Lie groups in their connection to geometry and topology will be presented in Chapter 11.

Notice that any groups with discrete topology is a zero-dimensional Lie group. Direct product of Lie groups also has natural Lie group structure. Thus in the structural theory of Lie groups interest in concentrated primarily on connected Lie groups. However *discrete subgroups* of connected Lie groups are of great interest.

Abelian Lie groups have rather simple structure. First, $\mathbb{R}^n$ with addition as the group operation is a Lie group. All its closed subgroups and factor-groups by closed subgroups are also Lie groups. Proofs of those facts will be given in Chapter 11. Now we consider natural examples.

EXAMPLE 4.10.2. Any linear subspace of $\mathbb{R}^n$ is a Lie subgroup isomorphic to $\mathbb{R}^k$ for some $k < n$.

The integer lattice $\mathbb{Z}^k \subset \mathbb{R}^k \subset R^n$ is a discrete subgroup and the factor group $\mathbb{R}^n/\mathbb{Z}^k$ is isomorphic to $\mathbb{T}^k \times \mathbb{R}^{n-k}$ and is a Lie group. In particular the torus $\mathbb{T}^n = \mathbb{R}^n/\mathbb{Z}^n$ is a compact connected abelian Lie group.

EXERCISE 4.10.1. Prove that the group $\mathbb{C}^*$ of non-zero complex numbers with multiplication as group operation is isomorphic to $\mathbb{R} \times \mathbb{S}^1$.

The group $GL(n, \mathbb{R})$ is the group of all invertible $n \times n$ matrices with differentiable structure inherited from its representation as the open subset of $\mathbb{R}^{n^2}$ determined by the condition $\det A \neq 0$ as in Example 4.1.8. Those groups play in the theory of Lie groups role somewhat similar to that played by the Euclidean spaces in the theory of differentiable manifolds. Many manifolds naturally appear as submanifolds of $\mathbb{R}^n$ and many more are diffeomorphic to submanifolds of $\mathbb{R}^n$ (see Theorem 4.5.1). The situation with Lie groups is similar. Most Lie groups naturally appear as Lie subgroups of $GL(n, \mathbb{R})$; such groups are called *linear groups*.

EXAMPLE 4.10.3. The *orthogonal group* $O(n)$ consists of all matrices A satisfying $AA^t = \mathrm{Id}$. Here the superscript t indicates transposition. It consists of two connected components according to the value of the determinant: +1 or -1. The former is also a group which is usually called the *special orthogonal group* and is denoted by $SO(n)$.

EXERCISE 4.10.2. Prove that $SO(2)$ is isomorphic to $\mathbb{S}^1$.

EXERCISE 4.10.3. Prove that $O(n)$ consist of matrices which represent all isometries of the Euclidean space $\mathbb{R}^n$ fixing the origin, or, equivalently, all isometries of the the unit sphere $\mathbb{S}^2 \subset \mathbb{R}^3$.

Many geometric structures naturally give rise to Lie groups, namely groups of transformations preserving the structure. In the example above the structure was the standard symmetric Riemannian metric on the sphere with $O(n)$ as the group of isometries. An even more basic example is given by $GL(n, \mathbb{R})$, the group of automorphisms of $\mathbb{R}^n$, the structure being that of linear space.

However, one needs to be cautious: this happens if for the group of transformations preserving the structure is *finite–dimensional*. For example, if one considers

$\mathbb{R}^n$ as a smooth manifold its automorphism group, the group of all diffeomorphisms, is not a Lie group.

EXERCISE 4.10.4. Identify isometries of the Euclidean plane with certain 3×3 matrices and prove that they form a linear group. Calculate its dimension.

For representation of groups of Euclidean isometries and affine transformations as linear groups see Exercise 4.11.14 and Exercise 4.11.15.

Notice that projective structure does not give as much new it terms of its group of automorphisms: projective transformations of $\mathbb{R}P(n)$ are simply linear transformations of $\mathbb{R}^{n+1}$. However, scalar matrices act identically on $\mathbb{R}P(n)$ so the group of projective transformations is not simply $GL(n+1,\mathbb{R})$ but its factor group.

If n is even and hence $n+1$ is odd one can find unique transformation with determinant one in each equivalence class, simply my multiplying all elements of a given matrix by the $(n+1)$ root of its determinant. hence in this case the group of projective transformations is isomorphic to $SL(n+1,\mathbb{R})$.

If n is odd the above procedure only works for matrices with positive determinant but it still leaves one non-identity matrix acting as identity, namely $-\operatorname{Id}$ which has determinant one in this case. On the other hand, matrices with negative determinant can be reduced to those with determinant -1, again with a similar identification. Thus the group of projective transformations in this case has a factor goup of index two which is isomorphic to $PSL(n+1,\mathbb{R}):=SL(n+1,\mathbb{R})/\{\pm\operatorname{Id}\}$.

EXAMPLE 4.10.4. The group $GL(n,\mathbb{C})$ of invertible $n\times n$ matrices with complex entries is a Lie group since it is an open subset $\det A\neq 0$ in the space of all $n\times n$ complex matrices which is isomorphic to $\mathbb{R}^{2n^2}$.

It is also a linear group since every complex number $a+bi$ can be identified with 2×2 real matrix $\begin{pmatrix} a & b\\ -b & a\end{pmatrix}$ and any $n\times n$ complex matrix can be associated with an $2n\times 2n$ real matrix by substituting each matrix element with the corresponding 2×2 matrix. This correspondence preserves addition and multiplication.

Its Lie subgroup $SL(n,\mathbb{C})$ consists of matrices with determinant one.

The group $GL(n,\mathbb{C})$ can be interpreted as the group of linear automorphisms of $\mathbb{R}^{2n}$ preserving and extra structure which in complex form corresponds to the multiplication of all coordinates of a vector by i.

EXAMPLE 4.10.5. The group $U(n)$ appears as groups of transformations of the space $\mathbb{C}^n$ preserving the *Hermitian product* $\sum_{i=1}^n z_i\bar{w}_i$ for $z=(z_1,\dots,z_n),\ w=(w_1,\dots,w_n)$. It is embedded into $GL(n,\mathbb{C})$ as the Lie subgroup of matrices A such that $AA^*\operatorname{Id}$. Here A^* is the matrix conjugate to A: its (i,j) matrix element is equal to the complex conjugate to the (j,i) element of A.

EXAMPLE 4.10.6. The symplectic group of $2n\times 2n$ consists of matrices A satisfying

$$AJA^t=J,\ \text{ where } J=\begin{pmatrix} 0 & \operatorname{Id}\\ -\operatorname{Id} & 0\end{pmatrix},$$

4.11. Problems

The next exercises are examples of smooth manifolds. Many examples of manifolds are given by configuration and phase spaces of mechanical systems. One can think of the configuration space of a mechanical system as a topological space whose points are different "positions" of the system, and neighborhoods are "nearby" positions (i.e., positions that can be obtained from the given one by motions of "length" smaller than a fixed number). The phase space of a mechanical system moving in time is obtained from its configuration space by supplying it with all possible velocity vectors.

EXERCISE 4.11.1. Describe the configuration space of the mechanical system consisting of a rod rotating in space about a fixed hinge at its extremity. What configuration space is obtained if the hinge is fixed at the midpoint of the rod?

EXERCISE 4.11.2. The *double pendulum* consists of two rods AB and CD moving in a vertical plane, connected by a hinge joining the extremities B and C, while the extremity A is fixed by a hinge in that plane. Find the configuration space of this mechanical system.

EXERCISE 4.11.3. On a round billiard table, a pointlike ball moves with uniform velocity, bouncing off the edge of the table according to the law saying that the angle of incidence is equal to the angle of reflection. Find the phase space of this system.

EXERCISE 4.11.4. Show that the configuration space of an asymetric solid rotating about a fixed hinge in 3-space is $\mathbb{R}P^3$.

EXERCISE 4.11.5. In $\mathbb{R}^9$ consider the set of points satisfying the following system of algebraic equations:

$$\begin{array}{ll} x_1^2+x_2^2+x_3^2=1; & x_1x_4+x_2x_5+x_3x_6=0; \\ x_4^2+x_5^2+x_6^2=1; & x_1x_7+x_2x_8+x_3x_9=0; \\ x_1^2+x_8^2+x_9^2=1; & x_4x_7+x_5x_8+x_6x_9=0. \end{array}$$

Show that this set is a smooth 3-dimensional submanifold of $\mathbb{R}^9$ and describe it. (Solution sets of systems of algebraic equations are not necessarily smooth manifolds: they may have *singularities*.)

EXERCISE 4.11.6. Show that the topological spaces obtained by identifying diametrically opposed points of the 3-sphere $\mathbb{S}^3$ and by identifying diametrically opposed boundary points of the 3-disk

$$\mathbb{D}^3 := \left\{(x_1,x_2,x_3)\in\mathbb{R}^3 \,|\, x_1^2+x_2^2+x_3^2\le 1\right\}$$

have a natural smooth manifold structure and are homeomorphic to each other.

EXERCISE 4.11.7. Seven rods of length 1 in the plane are joined end to end by hinges, and the two "free" ends are fixed to the plane by hinges at the distance 6.5 from each other. Find the configuration space of this mechanical system.

EXERCISE 4.11.8. Five rods of length 1 in the plane are joined end to end by hinges, and the two "free" ends are fixed to the plane by hinges at the distance 1 from each other. Find the configuration space of this mechanical system.

EXERCISE 4.11.9. Prove that the group $O(2)$ of orthogonal transformations of the plane is not isomorphic to $\mathbb{S}^1 \times C_2$.

ndard notation for cyclic groups

EXERCISE 4.11.10. Prove that the Lie group $SO(3)$ is diffeomorphic to the real projective space $\mathbb{R}P(3)$.

EXERCISE 4.11.11. Prove that the Lie group $SU(2)$ is diffeomorphic to the sphere $\mathbb{S}^3$.

EXERCISE 4.11.12. Represent the torus $\mathbb{T}^n$ as a linear group.

EXERCISE 4.11.13. What is the minimal value of m such that $\mathbb{T}^n$ is isomorphic to a Lie subgroup of $GL(m, \mathbb{R})$?

EXERCISE 4.11.14. Prove that the group of Euclidean isometries of of $\mathbb{R}^n$ is isomorphic to a Lie subgroup of $GL(n+1, \mathbb{R})$. Calculate its dimension.

EXERCISE 4.11.15. Prove that the group of affine transformations of $\mathbb{R}^n$ is isomorphic to a Lie subgroup of $GL(n+1, \mathbb{R})$. Calculate its dimension.

CHAPTER 5

TOPOLOGY AND GEOMETRY OF SURFACES

Compact (and some noncompact) surfaces are a favorite showcase for various branches of topology and geometry. They are two-dimensional topological manifolds, which can be supplied with a variety of naturally defined differentiable and Riemannian structures. Their complete topological classification, which coincides with their smooth (differentiable) classification, is obtained via certain simple invariants. These invariants allow a variety of interpretations: combinatorial, analytical and geometrical.

Surfaces are also one-dimensional complex manifolds; but, surprisingly, the complex stuctures are not all equivalent (except for the case of the sphere), although they can be classified. This classification if the first result in a rather deep area at the junction of analysis, geometry, and algebraic geometry known as *Teichmüller theory*, which recently has led to spectacular applications in theoretical physics.

In this chapter we study the classification of compact surfaces (two-dimensional manifolds) from various points of view. We start with a fundamental preparatory result, which we will prove by using a beautiful argument based on combinatorial considerations.

5.1. Two big separation theorems: Jordan and Schoenflies

The goal of this section is to prove the famous Jordan Curve Theorem, which we will need in the next section, and which is constantly used in many areas of analysis and topology. Note that although the statement of the theorem seems absolutely obvious, it does not have a simple proof.

5.1.1. Statement of the theorem and strategy of proof. Here we state the theorem and outline the main steps of the proof.

DEFINITION 5.1.1. A simple closed curve on a manifold M (in particular on the plane $\mathbb{R}^2$) is the homeomorphic image of the circle $\mathbb{S}^1$ in M, or equivalently the image of $\mathbb{S}^1$ under a topological embedding $\mathbb{S}^1 \to M$.

THEOREM 5.1.2 (Jordan Curve Theorem). *A simple closed curve C on the plane $\mathbb{R}^2$ separates the plane into two connected components.*

COROLLARY 5.1.3. *A simple closed curve C on the sphere $\mathbb{S}^2$ separates the sphere into two connected components.*

PROOF. The proof is carried out by a simple but clever reduction of the Jordan Curve Theorem to the nonplanarity of the graph $K_{3,3}$, established in ??

Suppose that C is an arbitrary (not necessarily polygonal) simple closed curve in the plane $\mathbb{R}^2$. Suppose l and m are parallel support lines of C and p is a line perpendicular to them and not intersecting the curve. Let A_1 and A_2 be points of the intersections of C with l and m, respectively. Further, let B_3 be the intersection point of l and p. The points A_1 and A_2 divide the curve C into two arcs, the "upper" one and the "lower" one. Take a line q in between l and m parallel to them. By compactness, there is a lowest intersection point B_1 of q with the upper arc and a highest intersection point B_2 of q with the lower arc. Let A_3 be an inner point of the segment $[B_1, B_2]$ (see the figure).

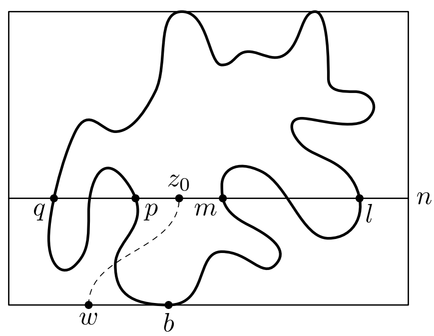

FIGURE 5.1.1. Proof of the Jordan Curve Theorem

We claim that $\mathbb{R}^2 \setminus C$ is not path connected, in fact there is no path joining A_3 and B_3. Indeed, if such a path existed, by Lemma **??** there would be an arc joining these two points. Then we would have nine pairwise nonintersecting arcs joining each of the points A_1, A_2, A_3 with all three of the points B_1, B_2, B_3. This means that we have obtained an embedding of the graph $K_{3,3}$ in the plane, which is impossible by Theorem 5.2.4. $\square$

5.1.2. Schoenflies Theorem. The Schoenflies Theorem is an addition to the Jordan curve theorem asserting that the curve actually bounds a disk. We state this theorem here without proof.

THEOREM 5.1.4 (Schoenflies Theorem). *A simple closed curve C on the plane $\mathbb{R}^2$ separates the plane into two connected components; the component with bounded closure is homeomorphic to the disk, that is,*

$$\mathbb{R}^2 \smallsetminus C = \mathcal{D}_1 \cup \mathcal{D}_2, \text{ where } \mathcal{D}_1 \cap \mathcal{D}_2 = \varnothing \text{ and } \overline{\mathcal{D}_1} \approx \mathbb{D}^2.$$

COROLLARY 5.1.5. *A simple closed curve C on the sphere $\mathbb{S}^2$ separates the sphere into two connected components, each of which has closure homeomorphic to the disk, that is,*

$$\mathbb{S}^2 \smallsetminus C = \mathcal{D}_1 \cup \mathcal{D}_2, \quad \text{where } \mathcal{D}_1 \cap \mathcal{D}_2 = \varnothing \text{ and } \overline{\mathcal{D}_i} \approx \mathbb{D}^2,\ i = 1, 2.$$

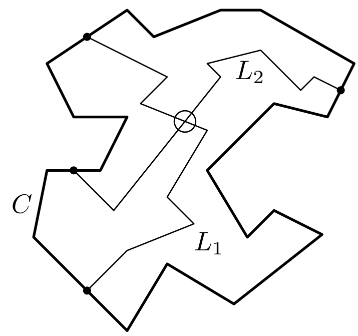

FIGURE 5.2.1. The polygonal lines L_1 and L_2 must intersect

5.2. Planar and non-planar graphs

5.2.1. Non-planarity of $K_{3,3}$. We first show that the graph $K_{3,3}$ has no polygonal embedding into the plane, and then show that it has no topological embedding in the plane.

PROPOSITION 5.2.1. [The Jordan curve theorem for broken lines] *Any broken line C in the plane without self-intersections splits the plane into two path connected components and is the boundary of each of them.*

PROOF. Let D be a small disk which C intersects along a line segment, and thus divides D into two (path) connected components. Let p be any point in $\mathbb{R}^2 \setminus C$. From p we can move along a polygonal line as close as we like to C and then, staying close to C, move inside D. We will then be in one of the two components of $D \setminus C$, which shows that $\mathbb{R}^2 \setminus C$ has *no more than two* components.

It remains to show that $\mathbb{R}^2 \setminus C$ is not path connected. Let ρ be a ray originating at the point $p \in \mathbb{R}^2 \setminus C$. The ray intersects C in a finite number of segments and isolated points. To each such point (or segment) assign the number 1 if C crosses ρ there and 0 if it stays on the same side. Consider the parity $\pi(p)$ of the sum S of all the assigned numbers: it changes continuously as ρ rotates and, being an integer, $\pi(p)$ is constant. Clearly, $\pi(p)$ does not change inside a connected component of $\mathbb{R}^2 \setminus C$. But if we take a segment intersecting C at a non-zero angle, then the parity π at its end points differs. This contradiction proves the proposition. □

We will call a closed broken line without self-intersections a *simple polygonal line*.

COROLLARY 5.2.2. *If two broken lines L_1 and L_2 without self-intersections lie in the same component of $\mathbb{R}^2 \setminus C$, where C is a simple closed polygonal line, with their endpoints on C in alternating order, then L_1 and L_2 intersect.*

PROOF. The endpoints a and c of L_1 divide the polygonal curve C into two polygonal arcs C_1 and C_2. The curve C and the line L_1 divide the plane into three path connected domains: one bounded by C, the other two bounded by the closed curves $C_i \cup L$, $i = 1, 2$ (this follows from Proposition 5.2.1). Choose points b and d on L_2 close to its endpoints. Then b and d must lie in different domains bounded by L_1 and C and any path joining them and not intersecting C, in particular L_2, must intersect L_1. □

PROPOSITION 5.2.3. *The graph $K_{3,3}$ cannot be polygonally embedded in the plane.*

PROOF. Let us number the vertices $x_1, \dots, x_6$ of $K_{3,3}$ so that its edges constitute a closed curve $C := x_1x_2x_3x_4x_5x_6$, the other edges being

$$E_1 := x_1x_4, \quad E_2 := x_2x_5, \quad E_3 := x_3x_6.$$

Then, if $K_{3,3}$ lies in the plane, it follows from Proposition 5.2.1 that C divides the plane into two components. One of the two components must contain at least two of the edges E_1, E_2, E_3, which then have to intersect (by Corollary 5.2.2). This is a contradiction which proves the proposition. □

THEOREM 5.2.4. *The graph $K_{3,3}$ is nonplanar, i.e., there is no topological embedding $h : K_{3,3} \hookrightarrow \mathbb{R}^2$.*

The theorem is an immediate consequence of the nonexistence of a PL-embedding of $K_{3,3}$ (Proposition 5.2.3) and the following lemma.

LEMMA 5.2.5. *If a graph G is planar, then there exists a polygonal embedding of G into the plane.*

PROOF. Given a graph $G \subset \mathbb{R}^2$, we first modify it in small disk neighborhoods of the vertices so that the intersection of (the modified graph) G with each disk is the union of a finite number of radii of this disk. Then, for each edge, we cover its complement to the vertex disks by disks disjoint from the other edges, choose a finite subcovering (by compactness) and, using the chosen disks, replace the edge by a polygonal line. □

5.2.2. Euler characteristic and Euler theorem. The *Euler characteristic* of a graph G without loops *embedded in the plane* is defined as

$$\chi(G) \ := \ V - E + F,$$

where V is the number of vertices and E is the number of edges of G, while F is the number of connected components of $\mathbb{R}^2 \setminus G$ (including the unbounded component).

THEOREM 5.2.6. [Euler Theorem] *For any connected graph G without loops embedded in the plane, $\chi(G) \ = \ 2$.*

PROOF. At the moment we are only able to prove this theorem for polygonal graphs. For the general case we will need Jordan curve Theorem Theorem 5.1.2. The proof will be by induction on the number of edges. For the graph with zero edges, we have $V = 1$, $E = 0$, $F = 1$, and the formula holds. Suppose it holds for all graphs with n edges; then it is valid for any connected subgraph H of G with n edges; take an edge e from G which is not in H but incident to H, and add it to H. Two cases are possible.

Case 1. Only one endpoint of e belongs to H. Then F is the same for G as for H and both V and E increase by one.

Case 2. Both endpoints of e belong to to H. Then e lies inside a face of H and divides it into two.[1] Thus by adding e we increase both E and F by one and leave V unchanged. Hence the Euler characteristic does not change. □

5.2.3. Kuratowski Theorem. We conclude this subsection with a beautiful theorem, which gives a simple geometrical obstruction to the planarity of graphs. We do not present the proof (which is not easy), because this theorem, unlike the previous one, is not used in the sequel.

small print for pa
the main line: no

THEOREM 5.2.7. [Kuratowski] *A graph is nonplanar if and only if it contains, as a topological subspace, the graph $K_{3,3}$ or the graph K_5.*

REMARK 5.2.8. The words "as a topological subspace" are essential in this theorem. They cannot be replaced by "as a subgraph": if we subdivide an edge of K_5 by adding a vertex at its midpoint, then we obtain a nonplanar graph that does *not* contain either $K_{3,3}$ or K_5.

EXERCISE 5.2.1. Can the graph $K_{3,3}$ be embedded in (a) the Möbius strip, (b) the torus?

EXERCISE 5.2.2. Is there a graph that cannot be embedded into the torus?

EXERCISE 5.2.3. Is there a graph that cannot be embedded into the Möbius strip?

5.3. Surfaces and their triangulations

In this section, we define (two-dimensional) surfaces, which are topological spaces that locally look like $\mathbb{R}^2$ (and so are supplied with local systems of coordinates). It can be shown that surfaces can always be triangulated (supplied with a PL-structure) and smoothed (supplied with a smooth manifold structure). We will not prove these two assertions here and limit ourselves to the study of triangulated surfaces (also known as two-dimensional PL-manifolds). The main result is a neat classification theorem, proved by means of some simple piecewise linear techniques and with the help of the Euler characteristic.

proof will be addec
an easy conse

[1] It is here that we need the conclusion of Jordan curve Theorem Theorem 5.1.2 in the case of general graphs. The rest of the argument remains the same as for polygonal graphs.

5.3.1. Definitions and examples.

DEFINITION 5.3.1. A *closed surface* is a compact connected 2-manifold (without boundary), i.e., a compact connected space each point of which has a neighborhood homeomorphic to the open 2-disk $\operatorname{Int}\mathbb{D}^2$. In the above definition, connectedness can be replaced by path connectedness without loss of generality (see ??)

A *surface with boundary* is a compact space each point of which has a neighborhood homeomorphic to the open 2-disk $\operatorname{Int}\mathbb{D}^2$ or to the open half disk

$$\operatorname{Int}\mathbb{D}^2_{1/2} = \{(x,y) \in \mathbb{R}^2 | x \geqslant 0, x^2 + y^2 < 1\}.$$

EXAMPLE 5.3.2. Familiar surfaces are the 2-sphere $\mathbb{S}^2$, the projective plane $\mathbb{R}P^2$, and the torus $\mathbb{T}^2 = S^1 \times S^1$, while the disk $\mathbb{D}^2$, the annulus, and the Möbius band are examples of surfaces with boundary.

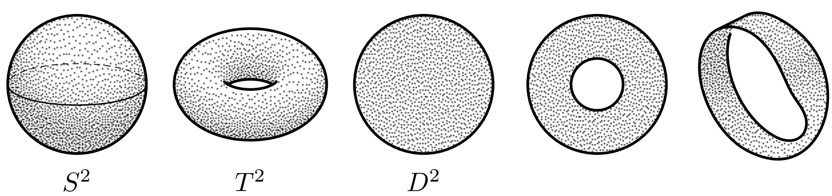

FIGURE 5.3.1. Examples of surfaces

DEFINITION 5.3.3. The *connected sum* $M_1 \# M_2$ of two surfaces M_1 and M_2 is obtained by making two small holes (i.e., removing small open disks) in the surfaces and gluing them along the boundaries of the holes

EXAMPLE 5.3.4. The connected sum of two projective planes $\mathbb{R}P^2 \# \mathbb{R}P^2$ is the famous *Klein bottle*, which can also be obtained by gluing two Möbius bands along their boundaries (see Fig.??). The connected sum of three tori $\mathbb{T}^2 \# \mathbb{T}^2 \# \mathbb{T}^2$ is (topologically) the surface of a pretzel (see Fig.??).

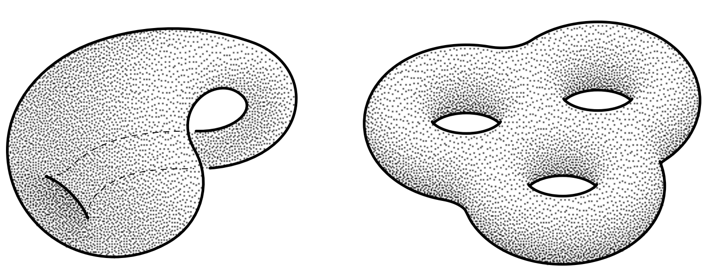

FIGURE 5.3.2. Klein bottle and pretzel

5.3.2. Polyhedra and triangulations. Our present goal is to introduce a combinatorial structure (called PL-structure) on surfaces. First we we give the corresponding definitions related to PL-structures.

A (finite) 2-*polyhedron* is a topological space represented as the (finite) union of triangles (its *faces* or *2-simplices*) so that the intersection of two triangles is either empty, or a common side, or a common vertex. The sides of the triangles are called *edges* or 1-*simplices*, the vertices of the triangles are called *vertices* or 0-*simplices* of the 2-polyhedron.

Let P be a 2-polyhedron and $v \in P$ be a vertex. The (closed) *star* of v in P (notation $\mathrm{Star}(v, P)$) is the set of all triangles with vertex v. The *link* of v in P (notation $\mathrm{Link}(v, P)$) is the set of sides opposite to v in the triangles containing v.

A finite 2-polyhedron is said to be a *closed PL-surface* (or a *closed triangulated surface*) if the star of any vertex v is homeomorphic to the closed 2-disk with v at the center (or, which is the same, if the links of all its vertices are homeomorphic to the circle).

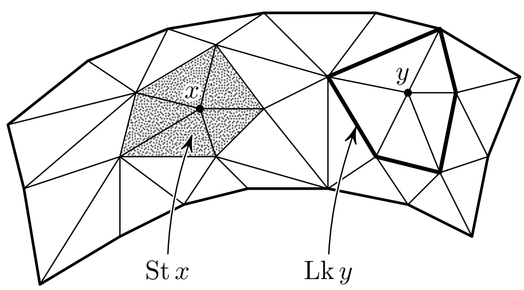

FIGURE 5.3.3. Star and link of a point on a surface

A finite 2-polyhedron is said to be a PL*-surface with boundary* if the star of any vertex v is homeomorphic either to the closed 2-disk with v at the center or to the closed disk with v on the boundary (or, which is the same, if the links of all its vertices are homeomorphic either to the circle or to the line segment). It is easy to see that in a PL-surface with boundary the points whose links are segments (they are called *boundary points*) constitute a finite number of circles (called *boundary circles*). It is also easy to see that each edge of a closed PL-surface (and each nonboundary edge of a surface with boundary) is contained in exactly two faces.

A PL-surface (closed or with boundary) is called *connected* if any two vertices can be joined by a sequence of edges (each edge has a common vertex with the previous one). Further, unless otherwise stated, we consider only connected PL-surfaces.

A PL-surface (closed or with boundary) is called *orientable* if its faces can be coherently oriented; this means that each face can be oriented (i.e., a cyclic order of its vertices chosen) so that each edge inherits opposite orientations from the orientations of the two faces containing this edge. An *orientation* of an orientable surface is a choice of a coherent orientation of its faces; it is easy to see that that any orientable (connected!) surface has exactly two orientations.

A *face subdivision* is the replacement of a face (triangle) by three new faces obtained by joining the baricenter of the triangle with its vertices. An *edge subdivision* is the replacement of the two faces (triangles) containing an edge by four new faces obtained by joining the midpoint of the edge with the two opposite vertices of the two triangles. A *baricentric subdivision* of a face is the replacement of a face (triangle) by six new faces obtained by constructing the three medians of the triangles. A *baricentric subdivision of a surface* is the result of the baricentric subdivision of all its faces. Clearly, any baricentric subdivision can be obtained by means of a finite number of edge and face subdivisions. A *subdivision* of a PL-surface is the result of a finite number of edge and face subdivisions.

Two PL-surfaces M_1 and M_2 are called *isomorphic* if there exists a homeomorphism $h : M_1 \to M_2$ such that each face of M_1 is mapped onto a face of M_2. Two PL-surfaces M_1 and M_2 are called *PL-homeomorphic* if they have isomorphic subdivisions.

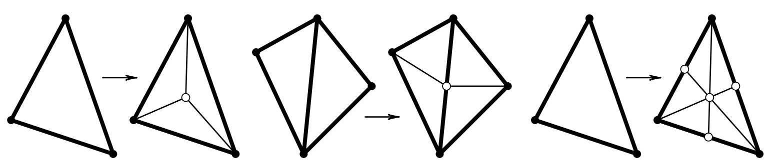

FIGURE 5.3.4. Face, edge, and baricentric subdivisions

EXAMPLE 5.3.5. Consider any convex polyhedron P; subdivide each of its faces into triangles by diagonals and project this radially to a sphere centered in any interior point of P. The result is a triangulation of the sphere.

If P is a tetrahedron the triangulation has four vertices. This is the minimal number of vertices in a triangulation of any surface. In fact, any triangulation of a surface with four vertices is equivalent of the triangulation obtained from a tetrahedron and thus for any surface other than the sphere the minimal number of vertices in a triangulation is greater then four.

EXERCISE 5.3.1. Prove that there exists a triangulation of the projective plane with any given number $N > 4$ of vertices.

EXERCISE 5.3.2. Prove that minimal number of vertices in a triangulation of the torus is six.

5.4. Euler characteristic and genus

In this section we introduce, in an elementary combinatorial way, one of the simplest and most important homological invariants of a surface M – its Euler characteristic $\chi(M)$. The Euler characteristic is an integer (actually defined for a much wider class of objects than surfaces) which is topologically invariant (and, in fact, also homotopy invariant). Therefore, if we find that two surfaces have different Euler characteristics, we can conclude that they are not homeomorphic.

5.4.1. Euler characteristic of polyhedra.

DEFINITION 5.4.1. The *Euler characteristic* $\chi(M)$ of a two-dimensional polyhedron, in particular of a PL-surface, is defined by

$$\boxed{\chi(M) := V - E + F},$$

where V, E, and F are the numbers of vertices, edges, and faces of M, respectively.

PROPOSITION 5.4.2. *The Euler characteristic of a surface does not depend on its triangulation. PL-homeomorphic PL-surfaces have the same Euler characteristic.*

PROOF. It follows from the definitions that we must only prove that the Euler characteristic does not change under subdivision, i.e., under face and edge subdivision. But these two facts are proved by a straightforward verification. □

EXERCISE 5.4.1. Compute the Euler characteristic of the 2-sphere, the 2-disk, the projective plane and the 2-torus.

EXERCISE 5.4.2. Prove that $\chi(M\#N) = \chi(M) + \chi(N) - 2$ for any PL-surfaces M and N. Use this fact to show that adding one handle to an oriented surface decreases its Euler characteristic by 2.

5.4.2. The genus of a surface. Now we will relate the Euler characteristic with a a very visual property of surfaces – their genus (or number of handles). The genus of an oriented surface is defined in the next section (see ??), where it will be proved that the genus g of such a surface determines the surface up to homeomorphism. The model of a surface of genus g is the *sphere with g handles*; for $g = 3$ it is shown on the figure.

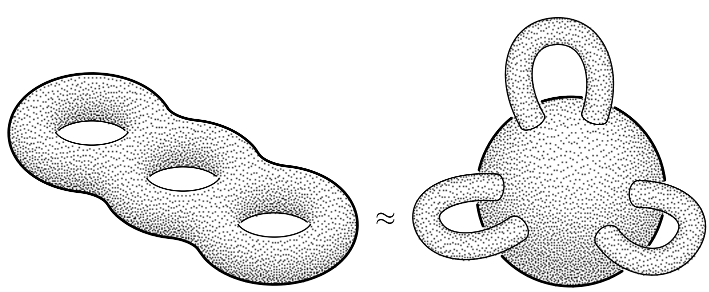

FIGURE 5.4.1. The sphere with three handles

PROPOSITION 5.4.3. *For any closed surface M, the genus $g(M)$ and the Euler characteristic $\chi(M)$ are related by the formula*

$$\boxed{\chi(M) \ = \ 2 - 2g(M)}.$$

PROOF. Let us prove the proposition by induction on g. For $g = 0$ (the sphere), we have $\chi(\mathbb{S}^2) = 2$ by Exercise **??**. It remains to show that adding one handle decreases the Euler characteristic by 2. But this follows from Exercise **??** $\square$

REMARK 5.4.4. In fact $\chi = \beta_2 - \beta_1 + \beta_0$, where the β_i are the Betti numbers (defined in **??**). For the surface of genus g, we have $\beta_0 = \beta_2 = 1$ and $\beta_1 = 2g$, so we do get $\chi = 2 - 2g$.

5.5. Classification of surfaces

In this section, we present the topological classification (which coincides with the combinatorial and smooth ones) of surfaces: closed orientable, closed nonorientable, and surfaces with boundary.

5.5.1. Orientable surfaces. The main result of this subsection is the following theorem.

THEOREM 5.5.1 (Classification of orientable surfaces). *Any closed orientable surface is homeomorphic to one of the surfaces in the following list*

$$\mathbb{S}^2,\ \mathbb{S}^1\times\mathbb{S}^1\ (\textit{torus}),\ (\mathbb{S}^1\times\mathbb{S}^1)\#(\mathbb{S}^1\times\mathbb{S}^1)\ (\textit{sphere with 2 handles}),\dots$$

$$\dots,(\mathbb{S}^1\times\mathbb{S}^1)\#(\mathbb{S}^1\times\mathbb{S}^1)\#\dots\#(\mathbb{S}^1\times\mathbb{S}^1)\ (\textit{sphere with k handles}),\dots$$

Any two surfaces in the list are not homeomorphic.

PROOF. By **??** we may assume that M is triangulated and take the double baricentric subdivision M'' of M. In this triangulation, the star of a vertex of M'' is called a *cap*, the union of all faces of M'' intersecting an edge of M but not contained in the caps is called a *strip*, and the connected components of the union of the remaining faces of M'' are called *patches*.

Consider the union of all the edges of M; this union is a graph (denoted G). Let G_0 be a maximal tree of G. Denote by M_0 the union of all caps and strips surrounding G_0. Clearly M_0 is homeomorphic to the 2-disk (why?). If we successively add the strips and patches from $M - M_0$ to M_0, obtaining an increasing sequence

$$M_0 \subset M_1 \subset M_2 \subset \cdots \subset M_p = M,$$

we shall recover M.

Let us see what happens when we go from M_0 to M_1.

If there are no strips left, then there must be a patch (topologically, a disk), which is attached along its boundary to the boundary circle Σ_0 of M_0; the result is a 2-sphere and the theorem is proved.

Suppose there are strips left. At least one of them, say S, is attached along one end to Σ_0 (because M is connected) and its other end is also attached to Σ_0 (otherwise S would have been part of M_0). Denote by K_0 the closed collar neighborhood of Σ_0 in M_0. The collar K_0 is homoeomorphic to the annulus (and not to the Möbius strip) because M is orientable. Attaching S to M_0 is the same as

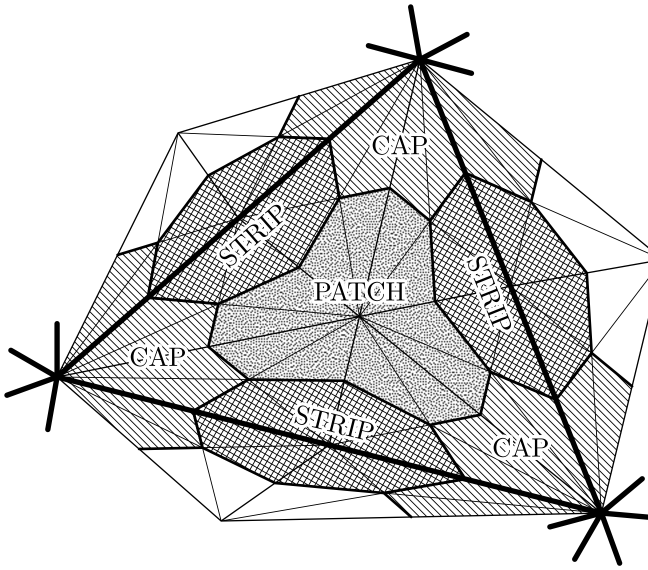

FIGURE 5.5.1. Caps, strips, and patches

attaching another copy of $K \cup S$ to M_0 (because the copy of K can be homeomorphically pushed into the collar K). But $K \cup S$ is homeomorphic to the disk with two holes (what we have called "pants"), because S has to be attached in the orientable way in view of the orientability of M (for that reason the twisting of the strip shown on the figure cannot occur). Thus M_1 is obtained from M_0 by attaching the pants $K \cup S$ by the waist, and M_1 has two boundary circles.

FIGURE ??? This cannot happen

Now let us see what happens when we pass from M_1 to M_2.

If there are no strips left, there are two patches that must be attached to the two boundary circles of M_1, and we get the 2-sphere again.

Suppose there are patches left. Pick one, say S, which is attached at one end to one of the boundary circles, say Σ_1 of M_1. Two cases are possible: either

(i) the second end of S is attached to Σ_2, or

(ii) the second end of S is attached to Σ_1.

Consider the first case. Take collar neighborhoods K_1 and K_2 of Σ_1 and Σ_2; both are homoeomorphic to the annulus (because M is orientable). Attaching S to M_1 is the same as attaching another copy of $K_1 \cup K_2 \cup S$ to M_1 (because the copy of $K_1 \cup K_2$ can be homeomorphically pushed into the collars K_1 and K_2).

FIGURE ??? Adding pants along the legs

But $K - 1 \cup K_2 \cup S$ is obviously homeomorphic to the disk with two holes. Thus, in the case considered, M_2 is obtained from M_1 by attaching pants to M_1 along the legs, thus decreasing the number of boundary circles by one,

The second case is quite similar to adding a strip to M_0 (see above), and results in attaching pants to M_1 along the waist, increasing the number of boundary circles by one.

What happens when we add a strip at the ith step? As we have seen above, two cases are possible: either the number of boundary circles of M_{i-1} increases by one or it decreases by one. We have seen that in the first case "inverted pants" are attached to M_{i-1} and in the second case "upright pants" are added to M_{i-1}.

FIGURE ??? Adding pants along the waist

After we have added all the strips, what will happen when we add the patches? The addition of each patch will "close" a pair of pants either at the "legs" or at the "waist". As the result, we obtain a sphere with k handles, $k \geqslant 0$. This proves the first part of the theorem.

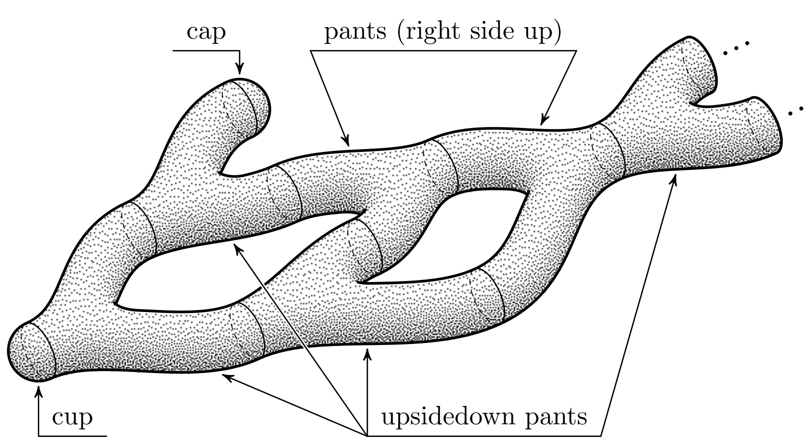

FIGURE 5.5.2. Constructing an orientable surface

To prove the second part, it suffices to compute the Euler characteristic (for some specific triangulation) of each entry in the list of surfaces (obtaining $2, 0, -2, -4, \ldots$, respectively). $\square$

5.5.2. Nonorientable surfaces and surfaces with boundary. Nonorientable surfaces are classified in a similar way. It is useful to begin with the best-known example, the *Möbius strip*, which is the nonorientable surface with boundary obtained by identifying two opposite sides of the unit square $[0,1]\times[0,1]$ via $(0,t)\sim(1,1-t)$. Its boundary is a circle.

Any compact nonorientable surface is obtained from the sphere by attaching several *Möbius caps*, that is, deleting a disk and identifying the resulting boundary circle with the boundary of a Möbius strip. Attaching m Möbius caps yields a surface of genus $2-m$. Alternatively one can replace any pair of Möbius caps by a handle, so long as at least one Möbius cap remains, that is, one may start from a sphere and attach one or two Möbius caps and then any number of handles.

All compact surfaces with boundary are obtained by deleting several disks from a closed surface. In general then a sphere with h handles, m Möbius strips, and d deleted disks has Euler characteristic

$$\chi = 2 - 2h - m - d.$$

In particular, here is the finite list of surfaces with nonnegative Euler characteristic:

Surface	h	m	d	χ	Orientable?
Sphere	0	0	0	2	yes
Projective plane	0	1	0	1	no
Disk	0	0	1	1	yes
Torus	1	0	0	0	yes
Klein bottle	0	2	0	0	no
Möbius strip	0	1	1	0	no
Cylinder	0	0	2	0	yes

5.6. The fundamental group of compact surfaces

Using the Seifert–van Kampen theorem (see ???), here we compute the fundamental groups of closed surfaces.

5.6.1. π_1 for orientable surfaces.

THEOREM 5.6.1. *The fundamental group of the orientable surface of genus g can be presented by $2g$ generators $p_1, m_1, \dots, p_n, m_n$ satisfying the following defining relation:*

$$p_1 m_1 p_1^{-1} m_1^{-1} \dots p_n m_n p_n^{-1} m_n^{-1} = 1.$$

PROOF. ++++++++++++++++++++++++++++++++++++ □

5.6.2. π_1 for nonorientable surfaces.

THEOREM 5.6.2. *The fundamental group of the nonorientable surface of genus g can be presented by the generators $c_1, \dots c_n$, where $n := 2g + 1$, satisfying the following defining relation:*

$$c_1^2 \dots c_n^2 = 1.$$

PROOF. ++++++++++++++++++++++++++++++++++ □

5.7. Vector fields on the plane

The notion of vector field comes from mechanics and physics. Examples: the velocity field of the particles of a moving liquid in hydrodynamics, or the field of gravitational forces in Newtonian mechanics, or the field of electromagnetic induction in electrodynamics. In all these cases, a vector is given at each point of some domain in space, and this vector changes continuously as we movefrom point to point.

In this section we will study, using the notion of degree (see??) a simpler model situation: vector fields on the plane (rather than in space).

5.7.1. Trajectories and singular points. A *vector field* V in the plane $\mathbb{R}^2$ is a rule that assigns to each point $p \in \mathbb{R}^2$ a vector $V(p)$ issuing from p. Such an assignment may be expressed in the coordinates x, y of $\mathbb{R}^2$ as

$$X = \alpha(x, y) \quad Y = \beta(x, y),$$

where $\alpha : \mathbb{R}^2 \to \mathbb{R}$ and $\beta : \mathbb{R}^2 \to \mathbb{R}$ are real-valued functions on the plane, (x, y) are the coordinates of the point p, and (X, Y) are the coordinates of the vector $V(p)$. If the functions α and β are continuous (respectively differentiable), then the vector field V is called *continuous* (resp. *smooth*).

A *trajectory* through the point $p \in \mathbb{R}^2$ is a curve $\gamma : \mathbb{R} \to \mathbb{R}^2$ passing through p and tangent at all its points to the vector field (i.e., the vector $V(q)$ is tangent to the curve $C := \gamma(\mathbb{R})$ at each point $q \in C$). A *singular point* p of a vector field V is a point where V vanishes: $V(p) = 0$; when V is a velocity field, such a point is often called a *rest point*, when V is a field of forces, it is called an *equilibrium point*.

5.7.2. Generic singular points of plane vector fields. We will now describe some of the simplestf singular points of plane vector fields. To define these points, we will not write explicit formulas for the vectors of the field, but instead describe the topological picture of its trajectories near the singular point and give physical examples of such singularities.

The *node* is a singular point contained in all the nearby trajectories; if all the trajectories move towards the point, the node is called *stable* and *unstable* if all the trajectories move away from the point. As an example, we can consider the gravitational force field of water droplets flowing down the surface $z = x^2 + y^2$ near the point $(0, 0, 0)$ (stable node) or down the surface $z = -x^2 - y^2$ near the same point (unstable node).

The *saddle* is a singular point contained in two transversal trajectories, called *separatrices*, one of which is ingoing, the other outgoing, the other trajectories behaving like a family of hyperbolas whose asymptotes are the separatrices. As an example, we can consider the gravitational force field of water droplets flowing down the surface $z = x^2 - y^2$ near the point $(0, 0, 0)$; here the separatrices are the coordinates axes.

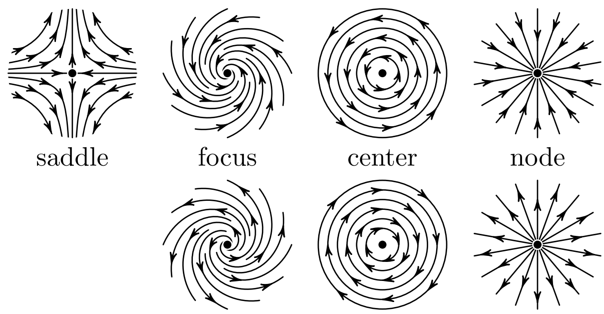

FIGURE 5.7.1. Simplest singular points of vector fields

The *focus* is a singular point that ressembles the node, except that the trajectories, instead of behaving like the set of straight lines passing through the point, behave as a family of logarithmic spirals converging to it (*stable focus*) or diverging from it (*unstable focus*).

topology of the noc
geometry is wror
"parabolas" t
horizontal line pl
and

The *center* is a singular point near which the trajectories behave like the family of concentric circles centered at that point; a center is called *positive* if the trajectories rotate counterclockwise and *negative* if they rotate clockwise. As an example, we can consider the velocity field obtained by rotating the plane about the origin with constant angular velocity.

REMARK 5.7.1. From the topological point of view, there is no difference between a node and a focus: we can unfurl a focus into a node by a homeomorphism which is the identity outside a small neighborhood of the singular point. However, we can't do this by means of a diffeomorphism, so that the node differs from the focus in the smooth category.

A singular point is called *generic* if it is of one of the first three types described above (node, saddle, focus). A vector field is called *generic* if it has a finite number of singular points all of which are generic. In what follows we will mostly consider generic vector fields.

REMARK 5.7.2. Let us explain informally why the term generic is used here. Generic fields are, in fact, the "most general" ones in the sense that, first, they occur "most often" (i.e., as close as we like to any vector field there is a generic one) and, second, they are "stable" (any vector field close enough to a generic one is also generic, has the same number of singular points, and those points are of the same types). Note that the center is *not* generic: a small perturbation transforms it into a focus. These statements are not needed in this course, so we will not make them more precise nor prove them.

REMARK 5.7.3. It can be proved that the saddle and the center are *not* topologically equivalent to each other and not equivalent to the node or to the focus; however, the focus and the node *are* topologically equivalent, as we noted above.

5.7.3. The index of plane vector fields. Suppose a vector field V in the plane is given. Let $\gamma : S^1 \to \mathbb{R}^2$ be a closed curve in the plane not passing through any singular points of V; denote $C := \gamma(S^1)$. To each vector $V(c)$, $c \in C$, let us assign the unit vector of the same direction as $V(c)$ issuing from the origin of coordinates $O \in \mathbb{R}^2$; we then obtain a map $g : C \to S^1_1$ (where $S^1_1 \subset \mathbb{R}^2$ denotes the unit circle centered at O), called the *Gauss map* corresponding to the vector field V and to the curve γ. Now we define the *index of the vector field* V *along the curve* γ as the degree of the Gauss map $g : S^1 \to S^1$ (for the definition of the degree of circle maps, see section 5, §3): $\operatorname{Ind}(\gamma, V) := \deg(g)$. Intuitively, the index is the total number of revolutions in the positive (counterclockwise) direction that the vector field performs when we go around the curve once.

REMARK 5.7.4. A simple way of computing $\operatorname{Ind}(\gamma)$ is to fix a ray issuing from O (say the half-axis Ox) and count the number of times p the endpoint of $V(c)$ passes through the ray in the positive direction and the number of times q in the negative one; then $\operatorname{Ind}(\gamma) = p - q$.

THEOREM 5.7.5. *Suppose that a simple closed curve γ does not pass through any singular points of a vector field V and bounds a domain that also does not contain any singular points of V. Then*

$$\boxed{\operatorname{Ind}(\gamma, V) = 0\ .}$$

PROOF. By the Schoenflies theorem, we can assume that there exists a homeomorphism of $\mathbb{R}^2$ that takes the domain bounded by $C := \gamma(\mathbb{S}^1)$ to the unit disk centered at the origin O. This homeomorphism maps the vector field V to a vector field that we denote by V'. Obviously,

$$\operatorname{Ind}(\gamma, V) = \operatorname{Ind}(S^1_0, V),$$

where S^1_O denotes the unit circle centered at O. Consider the family of all circles S^1_r of radius $r < 1$ centered at O. The vector $V'(O)$ is nonzero, hence for a small enough r_0 all the vectors $V'(s)$, $s \in S^1_{r_0}$, differ little in direction from $V'(O)$, so that $\operatorname{Ind}(S^1_r, V) = 0$. But then by continuity $\operatorname{Ind}(S^1_r, V) = 0$ for all $r \leqslant 1$. Now the theorem follows from (1). $\square$

Now suppose that V is a generic plane vector field and p is a singular point of V. Let C be a circle centered at p such that no other singular points are contained in the disk bounded by C. Then the *index of V at the singular point p* is defined as $\operatorname{Ind}(p, V) := \operatorname{Ind}(C, V)$. This index is well defined, i.e., it does not depend on the radius of the circle C (provided that the disk bounded by C does not contain any other singular points); this follows from the next theorem.

CHAPTER 6

COVERING SPACES AND DISCRETE GROUPS

We have already met covering spaces in Chapter 2, where our discussion was geometric and basically limited to definitions and examples. In this chapter, we return to this topic from a more algebraic point of view, which will allow us to produce numerous examples coming from group actions and to classify all covering spaces with given base (provided the latter is "nice" enough).

The main tools will be groups:

• discrete groups acting on manifolds, as the source of numerous examples of covering spaces;

• the fundamental groups of the spaces involved (and the homomorphisms induced by their maps), which will play the key role in the classification theorems of covering spaces.

6.1. Coverings associated with discrete group actions

We already mentioned (see ???) that one classical method for obtaining covering spaces is to consider discrete group actions on nice spaces (usually manifolds) and taking the quotient map to the orbit space. In this section, we dwell on this approach, providing numerous classical examples of covering spaces and conclude with a discussion of the generality of this method.

6.1.1. Discrete group actions. Here we present some basic definitions and facts related to group actions.

DEFINITION 6.1.1. The *action* of a group G on a set X is a map

$$G \times X \to X, \quad (g, x) \mapsto g(x),$$

such that

(1) $(gh)(x) = g(h(x))$;

(2) $e(x) = x$ if e is the unit element of G.

We will be interested in the case in which the group G is discrete (i.e., possesses, besides its group structure, the structure of a discrete topological space) and X is a Hausdorff topological space (more often than not a manifold). We then require the action to be a continuous map of topological spaces.

DEFINITION 6.1.2. The *orbit* of a point $x \in X$ is the set $\{g(x) \,|\, g \in G\}$. It immediately follows from definitions that the orbits constitute a partition of X. The *orbit space* of the action of G on X is the quotient space of X under the equivalence relation identifying all points lying in the same orbit; it is standardly denoted by X/G, and we have the *quotient map* $X \to X/G$ *of the action*.

We say that G *acts by homeomorphisms* if there is an inclusion (a monomorphism) $\mu : G \to \text{Homeo}(X)$ of the group G into the group Homeo(X) of homeomorphisms of X if μ satisfies $g(x) = (\mu(g))(x)$ for all $x \in X$ and all $g \in G$. This means that the assignment $(x, g) \mapsto g(x)$ is a homeomorphism of X. Similarily, for a metric space X, we say that a group G *acts by isometries* on X if there is a monomorphism $\mu : G \to \text{Isom}(X)$ of the group G into the group Isom(X) of isometries of X such that $g(x) = (\mu(g))(x)$ for all $x \in X$ and all $g \in G$. A similar meaning is assigned to the expressions *act by rotations*, *act by translations*, *act by homotheties*, etc.

Suppose G acts by homeomorphisms on a Hausdorff space X; we then say that the action of G is *normal* if any $x \in X$ has a neighborhood U such that all the images $g(U)$ for different $g \in G$ are disjoint, i.e.,

$$g_1(U) \cap g_2(U) \neq \varnothing \implies g_1 = g_2.$$

REMARK 6.1.3. There is no standard terminology for actions that we have called "normal"; sometimes the expressions "properly discontinuous action" or "covering space action" are used. We strongly favor "normal" – the reason for using it will become apparent in the next subsection.

EXAMPLE 6.1.4. Let X be the standard unit square centered at the origin of $\mathbb{R}^2$ and let G be its isometry group, acting on X in the natural way. Then the orbits of this action consist of 8, or 4, or 1 points (see the figure), the one-point orbit being the orbit of the origin. The orbit space can be visualized as a right isoceles triangle with the hypothenuse removed. Obviously, this action is *not* normal.

FIGURE ??? Isometry group acting on the square

EXAMPLE 6.1.5. Let the two-element group $\mathbb{Z}_2$ act on the 2-sphere $\mathbb{S}^2$ by symmetries with respect to its center. Then all the orbits consist of two points, the action is normal, and the orbit space is the projective plane $\mathbb{R}P^2$.

EXAMPLE 6.1.6. Let the permutation group S_3 act on the regular tetrahedron X by isometries. Then the orbits consist of 6, 4, or 1 point, and the action is not normal.

EXERCISE 6.1.1. Suppose X is the union of the boundary of the equilateral triangle and its circumscribed circle (i.e., the graph on 3 vertices of valency 3 and 6 edges). Find a normal group action for which the orbit space is the figure eight (i.e., the one-vertex graph with two edges (loops)).

EXERCISE 6.1.2. Find a normal action of the cyclic group $\mathbb{Z}_5$ on the annulus so that the orbit space is also the annulus.

6.1.2. Coverings as quotient maps to orbit spaces. The main contents of this subsection are its examples and exercises, which are all based on the following statement.

PROPOSITION 6.1.7. *The quotient map $X \to X/G$ of a Hausdorff topological space X to its orbit space X/G under a normal action* (*in the sense of Definition ???*) *of a discrete group G is a covering space .*

PROOF. By the definition of normal action, it follows that each orbit is in bijective correspondence with G (which plays the role of the fiber F in the definition of covering space). If x is any point of X and U is the neighborhood specified by the normality condition, each neighborhood of the family $\{g(U) \mid g \in G\}$ is projected homeomorphically onto another copy of U in the quotient space, which means that $p :\to X/G$ is a covering space. $\square$

REMARK 6.1.8. We will see later that in the situation of the proposition, the covering space will be normal (or regular, in another terminology), which means that the subgroup $p_{\#}(\pi_1(X)) \subset \pi_1(X/G)$ is normal. This explains our preference for the term "normal" for such group actions and such covering spaces.

EXAMPLE 6.1.9. Let the lattice $\mathbb{Z}^2$ act on the plane $\mathbb{R}^2$ in the natural way (i.e., by parallel translations along integer vectors). Then the orbit space of this action is the torus $\mathbb{T}^2$, and the corresponding quotient map $\tau : \mathbb{R}^2 \to \mathbb{T}^2$ is a covering. The covering map τ is actually a universal covering (i.e., it covers any other covering, see Definition ???), but we cannot prove this yet.

EXAMPLE 6.1.10. In the hyperbolic plane $\mathbb{H}^2$, choose a regular polygon P of $4g$ sides, $g \geq 2$, with inner angle $\pi/2g$. Consider the natural action of the subgroup G of isometries of $\mathbb{H}^2$ generated by parallel translations identifying opposite sides of the polygon. Then the entire hyperbolic plane will be covered by nonoverlapping copies of P.

Then the orbit space of this action is M_g^2, the sphere with g handles and the corresponding quotient map $\mu : \mathbb{H}^2 \to M_g^2$ is a covering. Since $\mathbb{H}^2$ is contractible, the map μ is the universal covering of M_g^2.

FIGURE ??? Universal covering of the sphere with two handles

EXAMPLE 6.1.11. Let X be the real line with identical little 2-spheres attached at its integer points. Let the group $\mathbb{Z}$ act on X by integer translations. The covering space corresponding to this action is represented in the figure.

FIGURE ??? Universal covering of the wedge of $\mathbb{S}^1$ and $\mathbb{S}^2$

EXERCISE 6.1.3. Construct a space X and an action of the cyclic group $\mathbb{Z}_3$ whose orbit space (i.e., the base of the corresponding covering space) is the wedge of the circle and the 2-sphere.

EXERCISE 6.1.4. Construct a 5-fold covering of the sphere with 11 handles over the sphere with 3 handles. Generalize to n-fold coverings (find k and l such that M_k is the n-fold cover of M_l)

6.1.3. Group actions and deck transformations. Recall that a deck transformation (see Definition ???) of a covering space $p : X \to B$ is an isomorphism of p to itself, i.e., a commutative diagram

$$\begin{array}{ccc} X & \xrightarrow{F} & X \\ \Big\downarrow p & & \Big\downarrow p \\ B & \xrightarrow{\mathrm{Id}} & B \end{array}$$

PROPOSITION 6.1.12. *The group of deck transformations of the covering space obtained as the quotient map of a Hausdorff space X to its orbit space X/G under a normal action* (*see Definition ???*) *of a discrete group G is isomorphic to the group G.*

PROOF. This immediately follows from the definition of normal action: the fiber of the covering space under consideration is an orbit of the action of G, and property (1) of the definition of an action (see ???) implies that the deck transformations can be identified with G. $\square$

6.1.4. Subgroup actions and associated morphisms. If G is a group acting by homeomorphisms on a Hausdorff space X, and H is a subgroup of G, then H acts by homeomorphisms on X in the obvious way.

PROPOSITION 6.1.13. *A subgroup H of a discrete group G possessing a normal action on a Hausdorff space X induces an injective morphism of the covering space p_H corresponding to H into the covering space p_G corresponding to G.*

In this case the image of p_H in p_G is called a *subcovering* of p_G.

PROOF. The statement of the proposition is an immediate consequence of definitions. $\square$

EXAMPLE 6.1.14. Let $w_n : \mathbb{S}^1 \to \mathbb{S}^1$ be the n-fold covering of the cicle by another copy of the circle, where $n = pq$ with p and q coprime. We can then consider two more coverings of the circle by itself, namely the p- and q-fold coverings $w_p,\ w_q : \mathbb{S}^1 \to \mathbb{S}^1$. Both of them are subcoverings of w_n, and their composition is (isomorphic to) w_n.

EXERCISE 6.1.5. Prove that the p-fold covering of the circle by itself has no subcoverings (other than the identity map) if p is prime.

EXERCISE 6.1.6. Describe (up to isomorphism) all the subcoverings of the universal covering of the torus.

6.2. The hierarchy of coverings, universal coverings

In this section we are interested in the "social life" of covering spaces, i.e., in how they interact with each other. In this study, a number of natural questions arise, for example concerning the "hierarchy" of coverings $p : X \to B$ over a fixed space B and the so-called universal coverings. These questions will mainly be answered in the next sections, but here the reader will find many useful constructions and examples.

6.2.1. Definitions and examples. Recall that the definition of covering space $p : X \to B$ implies that p is a local homeomorphism, and the number of points of $p^{-1}(b)$ (which can be finite or countable) does not depend on the choice of $b \in B$. Recall further that covering spaces form a category, morphisms being pairs of maps $f : B \to B'$, $F : X \to X'$ for which $f \circ p = p' \circ F$, i.e., the square diagram

$$\begin{array}{ccc} X & \xrightarrow{F} & X' \\ \Big\downarrow{\scriptstyle p} & & \Big\downarrow{\scriptstyle p'} \\ B & \xrightarrow{f} & B' \end{array}$$

is commutative. If the maps f and F are homeomorphisms, then the two covering spaces are isomorphic. We do not distinguish isomorphic covering spaces: the classification of covering spaces will always be performed up to isomorphism.

Consider a fixed Hausdorff topological space B and the set (which is actually a category) of covering spaces with base B. Our aim is to define an order in this set.

DEFINITION 6.2.1. We say that the covering space $p' : X' \to B$ *supersedes* or *covers* the covering space $p : X \to B$ (and write $p' \gg p$) if there is a covering space $q : X' \to X$ for which the following diagram

$$\begin{array}{ccc} X & \xrightarrow{q} & X' \\ \Big\downarrow{\scriptstyle p} & & \Big\downarrow{\scriptstyle p} \\ B & \xrightarrow{\mathrm{Id}} & B' \end{array}$$

is commutative.

The relation $\gg$ is obviously reflexive and transitive, so it is a partial order relation. Thus we obtain a hierarchy of covering spaces over each fixed base.

EXAMPLE 6.2.2. Let $w_n : \mathbb{S}^1 \to \mathbb{S}_1$ denote the n-fold covering of the circle by itself. Then w_m supersedes w_n iff n divides m. Further, the least (in the sense of the order $\gg$) covering w_d that supersedes both w_r and w_s is w_m, where m is the least common multiple of r and s, while the greatest covering superseded by both w_r and w_s is w_d, where d is the greatest commom divisor of r and s.

EXERCISE 6.2.1. Prove that no covering of the projective plane $\mathbb{R}P^2$ supersedes its double covering by the sphere $\mathbb{S}^2$.

EXERCISE 6.2.2. Prove that no covering of the circle $\mathbb{S}^1$ supersedes its covering by $\mathbb{R}$ (via the exponential map).

6.2.2. Universal coverings. Recall that the universal covering space $s : E \to B$ of a given space B was defined as a covering satisfying the condition $\pi_1(E) = 0$. It turns out that such a covering $s : E \to B$ exists, is unique (up to isomorphism), and coincides with the maximal covering of B with respect to the relation $\gg$, provided B is nice enough (if it has no local pathology, e.g. is a manifold, a simplicial space, or a CW-space). We could prove this directly now, but instead we will prove (later in this chapter) a more general result from which the above statements follow.

In this subsection, using only the definition of universal covering (the condition $\pi_1(E) = 0$), we will accumulate some more examples of universal coverings. Recall that we already know several: $\mathbb{R}$ over $\mathbb{S}^1$, $\mathbb{S}^2$ over $\mathbb{R}P^2$, $\mathbb{R}^2$ over $\mathbb{T}^2$.

EXAMPLE 6.2.3. The universal covering of the sphere $\mathbb{S}^n$, $n \geq 2$, and more generally of any simply connected space, is the identity map.

EXAMPLE 6.2.4. The universal covering of the wedge sum $\mathbb{S}^1 \vee \mathbb{S}^1$ of two circles is the infinite 4-valent graph Γ (shown on the figure) mapped onto $\mathbb{S}^1 \vee \mathbb{S}^1$ in the following way.

FIGURE ??? Universal covering of the wedge of $\mathbb{S}^1$ and $\mathbb{S}^1$

Each vertex of Γ is sent to the point of tangency of the wedge and each edge is wrapped once around one of the circles (you can think of this as an infinite process, beginning at the center of the graph and moving outwards in a uniform way). It follows that we have obtained the universal covering of $\mathbb{S}^1 \vee \mathbb{S}^1$, since Γ is (obviously) simply connected and the map described above is indeed a covering with fiber $\mathbb{Z}$ (this can be checked directly in the two different types of points of $\mathbb{S}^1 \vee \mathbb{S}^1$ – the tangency point, and all the others).

EXERCISE 6.2.3. Find an appropriate group G and define a normal action of G on the graph Γ from the previous example so as to obtain the universal covering of the wedge sum of two circles.

EXERCISE 6.2.4. Describe the universal covering of the wedge sum of three circles.

EXERCISE 6.2.5. Describe the universal covering of the union of three circles, two of which are tangent (at different points) to the third.

6.3. Path lifting and covering homotopy properties

In this section, we prove two important technical assertions which allow, given a covering space $p : X \to B$, to lift "upstairs" (i.e., to X) continuous processes taking place "downstairs" (i.e., in B). The underlying idea has already been exploited when we defined the degree of circle maps by using the exponential map (see ???), and we will now be generalizing the setting from the exponential map to arbitrary covering spaces.

6.3.1. Path lifting. Let $p : X \to B$ be a covering space. Recall that the *lift* of a map $f : A \to B$ was defined (see ???) as any map $\widetilde{f} : A \to X$ such that $p \circ \widetilde{f} = f$.

LEMMA 6.3.1 (Path lifting lemma). *Any path in the base of a covering space can be lifted to the covering, and the lift is unique if its initial point in the covering is specified. More precisely, if $p : X \to B$ is a covering space, $\alpha : [0,1] \to B$ is any path, and $x_0 \in p^{-1}(\alpha(0))$, then there exists a unique map $\widetilde{\alpha} : [0,1] \to X$ such that $p \circ \widetilde{\alpha} = \alpha$ and $\widetilde{\alpha}(0) = x_0$.*

PROOF. By the definition of covering space, for each point $b \in \alpha([0,1])$ there is a neighborhood U_b whose inverse image under p falls apart into disjoint neighborhoods each of which is projected homeomorphically by p onto U_b. The set of all such U_b covers $\alpha([0,1])$ and, since $\alpha([0,1])$ is compact, it possesses a finite subcover that we denote by $U_0, U_1, \dots U_k$.

Without loss of generality, we assume that U_0 contains $b_0 := \alpha(0)$ and denote by $\widetilde{U}_0$ the component of $p^{-1}(U_0)$ that contains the point x_0. Then we can lift a part of the path α contained in U_0 to $\widetilde{U}_0$ (uniquely!) by means of the inverse to the homeomorphism between $\widetilde{U}_0$ and U_0.

Now, again without loss of generality, we assume that U_1 intersects U_0 and contains points of $\alpha[0,1]$ not lying in U_0. Let $b_1 \in \alpha([0,1])$ be a point contained both in U_0 and U_1 and denote by $\widetilde{b}_1$ the image of b_1 under $p^{-1}\big|_{U_0}$. Let $\widetilde{U}_1$ be the component of the inverse image of U_1 containing $\widetilde{b}_1$. We now extend the lift of our path to its part contained in U_1 by using the inverse of the homeomorphism between $\widetilde{U}_1$ and U_1. Note that the lift obtained is the only possible one. Our construction is schematically shown on the figure.

FIGURE ??? Path lifting construction

Continuing in this way, after a finite number of steps we will have lifted the entire path $\alpha([0,1])$ to X, and the lift obtained will be the only one obeying the conditions of the lemma.

To complete the proof, it remains to show that the lift that we have constructed is unique and continuous. We postpone the details of this argument (which uses the fact that X and B are "locally nice", e.g. CW-spaces) to Subsection ??? . $\square$

REMARK 6.3.2. Note that the lift of a closed path is *not* necessarily a closed path, as we have already seen in our discussion of the degree of circle maps.

Note that if all paths (i.e., maps of $A = [0,1]$) can be lifted, it is not true that all maps of *any* space A can be lifted.

EXAMPLE 6.3.3. Let $\tau : \mathbb{R}^2 \to \mathbb{T}^2$ be the standard covering of the torus by the plane. Then the map $\alpha : \mathbb{S}^1 \to \mathbb{T}^2$ taking the circle to some meridian of the torus cannot be lifted to the plane. Indeed, if such a lift existed, it would be a continuous map of a compact set ($\mathbb{S}^1$) with a noncompact image.

EXERCISE 6.3.1. Give an example of a map $\alpha : A \to \mathbb{R}P^2$ which cannot be lifted to the standard covering space $p : \mathbb{S}^2 \to \mathbb{R}P^2$.

6.3.2. Homotopy lifting. Now we generalize the path lifting lemma to homotopies, having in mind that a path is actually a homotopy, namely a homotopy of the one-point space. This trivial observation is not only the starting point of the formulation of the covering homotopy theorem, but also the key argument in its proof.

THEOREM 6.3.4 (Covering homotopy theorem). *Any homotopy in the base of a covering space can be lifted to the covering, and the homotopy is unique if its initial map in the covering is specified as a lift of the initial map of the given homotopy. More precisely, if* $p : X \to B$ *is a covering,* $F : A \times [0,1] \to B$ *is any homotopy whose initial map* $f_0(\cdot) := F(\cdot, 0)$ *possesses a lift* $\widetilde{f}_0$*, then there exists a unique homotopy* $\widetilde{F} : A \times [0,1] \to X$ *such that* $p \circ \widetilde{F} = F$ *and* $\widetilde{F}(\cdot, 0) = \widetilde{f}_0(\cdot)$.

PROOF. The theorem will be proved by means of a beautiful trick, magically reducing the theorem to the path lifting lemma from the previous subsection. Fix some point $a \in A$. Define $\alpha_a(t) := F(a,t)$ and denote by x_a the point $\widetilde{f}_0(a)$. Then α_a is a path, and by the path lifting lemma, there exists a unique lift $\widetilde{\alpha}_a$ of this path such that $\widetilde{\alpha}(0) = x_a$. Now consider the homotopy defined by

$$\widetilde{F}(a,t) := \widetilde{\alpha}_a(t), \quad \text{for all} \quad a \in A, \quad t \in [0,1].$$

Then, we claim that $\widetilde{F}$ satisfies all the conditions of the theorem.

To complete the proof, one must verify that $\widetilde{F}$ is continuous and unique. We leave this verification to the reader. □

REMARK 6.3.5. Although the statement of the theorem is rather technical, the underlying idea is of fundamental importance. The *covering homotopy property* that it asserts holds not only for covering spaces, but more generally for arbitrary fiber bundles. Still more generally, this property holds for a very important class of fibrations, known as Serre fibrations (see ???), which are defined as precisely those which enjoy the covering homotopy property.

EXAMPLE 6.3.6. Let X be the union of the lateral surface of the cone and the half-line issuing from its vertex v, and let $p : X \to B$ be the natural projection of X on the line $B = \mathbb{R}$ (see the figure).

FIGURE ??? A map without the covering homotopy property

Then the covering homotopy property does not hold for p. Indeed, the path lifting property already fails for paths issuing from $p(v)$ and moving to the left of $p(v)$ (i.e., under the cone). Of course, lifts of such paths exist, but they are not unique, since they can wind around the cone in different ways.

6.4. Classification of coverings with given base via π_1

As we know, a covering space $p : X \to B$ induces a homomorphism $p_\# : \pi_1(X) \to \pi_1(B)$ (see ???). We will see that when the spaces X and B are "locally nice", $p_\#$ entirely determines (up to isomorphism) the covering space p over a given B.

More precisely, in this section we will show that, provided that the "local nicety" condition holds, $p_{\#}$ is a monomorphism and that, given a subgroup G of $\pi_1(B)$, we can effectively construct a unique space X and a unique (up to isomorphism) covering map $p : X \to B$ for which G is the image of $\pi_1(X)$ under $p_{\#}$. Moreover, we will prove that there is a bijection between conjugacy classes of subgroups of $\pi_1(B)$ and isomorphism classes of coverings, thus achieving the classification of all coverings over a given base B in terms of $\pi_1(B)$.

Note that here we are *not* assuming that G is a normal subgroup of $\pi_1(B)$, and so the covering space is not necessarily normal and G does not necessarily coincide with the group of deck transformations.

6.4.1. Injectivity of the induced homomorphism. The goal of this subsection is to prove the following theorem.

THEOREM 6.4.1. *The homomorphism $p_{\#} : \pi_1(X) \to \pi_1(B)$ induced by any covering space $p : X \to B$ is a monomorphism.*

PROOF. The theorem is an immediate consequence of the homotopy lifting property proved in the previous section. Indeed, it suffices to prove that a nonzero element $[\alpha]$ of $\pi_1(X)$ cannot be taken to zero by $p_{\#}$. Assume that $p_{\#}([\alpha]) = 0$. This means that the loop $p \circ \alpha$, where $\alpha \in [\alpha]$, is homotopic to a point in B. By the homotopy lifting theorem, we can lift this homotopy to X, which means that $[\alpha] = 0$. □

6.4.2. Constructing the covering space. Here we describe the main construction of this chapter: given a space and a subgroup of its fundamental group, we construct the associated covering. This construction works provided the space considered is "locally nice" in a sense that will be specified in the next subsection, and we will postpone the conclusion of the proof of the theorem until then.

THEOREM 6.4.2. *For any "locally nice" space B and any subgroup $G \subset \pi_1(B, b_0)$ there exists a unique covering space $p : X \to B$ such that $p_{\#}(X) = G$.*

PROOF. The theorem is proved by means of another magical trick. Let us consider the set $P(B, b_0)$ of all paths in B issuing from b_0. Two such paths $\alpha_i : [0, 1] \to B$, $i = 1, 2$ will be identified (notation $\alpha_1 \sim \alpha_2$) if they have a common endpoint and the loop λ given by

$$\lambda(t) = \begin{cases} \alpha_1(2t) & \text{if } 0 \le t \le 1/2, \\ \alpha_2(2 - 2t) & \text{if } 1/2 \le t \le 1. \end{cases}$$

determines an element of $\pi_1(B)$ that belongs to G. (The loop λ can be described as first going along α_1 (at double speed) and then along α_2 from its endpoint back to b_0, also at double speed.)

Denote by $X := P(B, b_0)/_{\sim}$ the identification space of $P(B, b_0)$ by the equivalence relation just defined. Endow X with the natural topology (the detailed definition appears in the next subsection) and define the map $p : X \to B$ by stipulating that it takes each equivalence class of paths in $P(B, b_0)$ to the endpoint of one of them (there is no ambiguity in this definition because equivalent paths have the same endpoint).

Then $p : X \to B$ is the required covering space. It remains to prove that

(i) p is continuous;

(ii) p is a local homeomorphism;

(iii) $p_{\#}(\pi_1(X))$ coincides with G;

(iv) p is unique.

This will be done in the next subsection. □

To better understand the proof, we suggest that the reader do the following exercise.

EXERCISE 6.4.1. Prove the theorem in the particular case $G = 0$, i.e., construct the universal covering of B,

6.4.3. Proof of continuity and uniqueness. In this subsection, we fill in the missing details of the previous exposition: we specify what is meant by "locally nice" and use that notion to prove the continuity of different key maps constructed above and give rigorous proofs of their uniqueness properties.

DEFINITION 6.4.3. A topological space X is called *locally path connected* if for any point $x \in X$ and any neighborhood U of x there exists a smaller neighborhood $V \subset U$ of x which is path connected. A topological space X is called *locally simply connected* if for any point $x \in X$ and any neighborhood U of x there exists a smaller neighborhood $V \subset U$ of x which is simply connected.

EXAMPLE 6.4.4. Let $X \subset \mathbb{R}^2$ be the union of the segments

$$\{(x, y) \mid y = 1/n,\ 0 \le x \le 1\} \quad n = 1, 2, 3, \dots$$

and the two unit segments $[0, 1]$ of the x-axis and y-axis (see the figure). Then X is path connected but not locally path connected (at all points of the interval $(0, 1]$ of the x-axis).

FIGURE ??? Not locally connected and not simply connected spaces

EXAMPLE 6.4.5. Let $X \subset \mathbb{R}^2$ be the union of the circles

$$\{(x,y) \mid x^2 + (y - 1/n)^2 = 1/n^2\} \quad n = 1, 2, 3, \dots;$$

the circles are all tangent to the x-axis and to each other at the point $(0,0)$ (see the figure). Then X is path connected but not locally simply connected (at the point$(0,0)$).

We will now conclude, step by step, the proof of the main theorem of the previous subsection under the assumption that B is locally path connected and locally simply connected.

(o) *Definition of the topology in* $X = P(B, b_0)/_\sim$. In order to define the topology, we will specify a base of open sets of rather special form, which will be very convenient for our further considerations. Let U be an open set in B and $x \in X$ be a point such that $p(x) \in U$. Let α be one of the paths in x with initial point x_0 and endpoint x_1. Denote by (U, x) the set of equivalence classes (with respect to $\sim$) of extensions of the path α whose segments beyond x_1 lie entirely inside U. Clearly, (U, x) does not depend on the choice of $\alpha \in x$.

We claim that (U, x) actually does not depend on the choice of the point x in the following sense: if $x_2 \in (U, x_1)$, then $(U, x_1) = (U, x_2)$. To prove this, consider the points $b_1 := p(x_1)$ and $b_2 := p(x_2)$. Join the points b_1 and (b_2) by a path (denoted β) contained in U (see the figure).

FIGURE ??? Defining the topology in the covering space

Let $\alpha\alpha_1$ denote an extension of α, with the added path segment α_1 contained in U. Now consider the path $\alpha\beta\beta^{-1}\alpha_1$, which is obviously homotopic to $\alpha\alpha_1$. On the other hand, it may be regarded as the extension (beyond x_2) of the path $\alpha\beta$ by the path $\beta^{-1}\alpha_{-1}$. Therefore, the assignment $\alpha\alpha_1 \mapsto \alpha\beta\beta^{-1}\alpha_1$ determines a bijection between (U, x_1) and (U, x_2), which proves our claim.

Now we can define the topology in X by taking for a base of the topology the family of all sets of the form (U, x). To prove that this defines a topology, we must check that a nonempty intersection of two elements of the base contains an element of the base. Let the point x belong to the intersection of the sets (U_1, x_1) and (U_2, x_2). Denote $V := U_1 \cap U_2$ and consider the set (V, x); this set is contained in the intersection of the sets (U_1, x_1) and (U_2, x_2) (in fact, coincides with it) and contains x, so that $\{(U, x)\}$ is indeed a base of a topology on X.

(i) *The map* p *is continuous.* Take $x \in X$. Let U be any path connected and simply connected neighborhood of $p(x)$ (it exists by the condition imposed on

B). The inverse image of U under p is consists of basis open sets of the topology of X (see item (o)) and is therefore open, which establishes the continuity at the (arbitrary) point $x \in X$.

(ii) *The map p is a local homeomorphism.* Take any point $x \in X$ and denote by $p|_U : (U, x) \to U$ the restriction of p to any basis neighborhood (U, x) of x, so that U will be an open path connected and simply connected set in B. The path connectedness of U implies the surjectivity of $p|_U$ and its simple connectedness, the injectivity of $p|_U$.

(iii) *The subgroup $p_\#(\pi_1(X))$ coincides with G.* Let α be a loop in B with basepoint b_0 and $\widetilde{\alpha}$ be the lift of α initiating at x_0 ($\widetilde{\alpha}$ is not necessarily a closed path). The subgroup $p_\#(\pi_1)(X)$ consists of homotopy classes of the loops α whose lifts $\widetilde{\alpha}$ are closed paths. By construction, the path $\widetilde{\alpha}$ is closed iff the equivalence class of the loop α corresponds to the point x_0, i.e., if the homotopy class of α is an element of G.

(iv) *The map p is unique.* To prove this we will need the following lemma.

LEMMA 6.4.6 (Map Lifting Lemma). *Suppose $p : X \to B$ is a covering space, $f : A \to B$ is a (continuous) map of a path connected and locally path connected spaces A and B, and $f_\#$ is a monomorphism of $\pi_1(A, a_0)$ into $p_\#(\pi_1(X, x_0))$. Then there exists a unique lift $\widetilde{f}$ of the map f, i.e., a unique map $\widetilde{f} : A \to X$ satisfying $p \circ \widetilde{f} = f$ and $\widetilde{f}(a_0) = x_0$.*

PROOF. Consider an arbitrary path α in A joining a_0 to some point a. The map f takes it to to the path $f \circ \alpha$. By the Path Lifting Lemma (see ???), we can lift $f \circ \alpha$ to a (unique) path $\widetilde{a}$ in X issuing from the point x_0. Let us define $\widetilde{f} : A \to X$ by setting $\widetilde{f}(a) = x$, where x is the endpoint of the path $\widetilde{\alpha}$.

First let us prove that $\widetilde{f}$ is well defined, i.e., does not depend on the choice of the path α. Let α_i, $i = 1, 2$, be two paths joining a_0 to a. Denote by λ the loop at a_0 defined by

$$\lambda(t) = \begin{cases} \alpha_1(2t) & \text{if } 0 \le t \le 1/2, \\ \alpha_2(2 - 2t) & \text{if } 1/2 \le t \le 1. \end{cases}$$

Then the lift of the loop $f \circ \lambda$ issuing from x_0 should be a closed path in X_0, i.e., the class of the loop $f \circ \alpha$ should lie in $p_\#\pi_1(X, x_0)$, i.e., we should have $f_\#\pi_1(A, a_0) \subset p_\#\pi_1(X, x_0)$. But this holds by assumption. Thus $\widetilde{f}$ is well defined.

It remains to prove that $\widetilde{f}$ is continuous. Let $a \in A$ and $x := \widetilde{f}(a)$. For the point $p(x)$, let us choose a path connected neighborhood U from the definition of covering space. Let $\widetilde{U}$ be the path connected component of $p^{-1}(U)$ containing the point x_0. Since f is continuous, $f^{-1}(U)$ contains a certain neighborhood of the point a. Since A is locally path connected, we can assume that V is path connected. Now we claim that $\widetilde{f}(V) \subset \widetilde{U}$ (which means that $\widetilde{f}$ is continuous). Indeed, any point $a_1 \in V$ can be joined to α_0 by a path α entirely contained in V. Its image

$\alpha \circ f$ lies in U, therefore $\alpha \circ f$ can be lifted to a path entirely contained in $\widetilde{U}$. But this means that $\widetilde{f}(a) \in \widetilde{U}$. $\square$

Now to prove the uniqueness of p (the covering space corresponding to the given subgroup of $\pi_1(B)$, where B is path connected and locally path connected), suppose that we have two coverings $p_i : X_i \to B,\ i = 1, 2$ such that

$$(p_1)_\#\big(\pi_1(X-1, x_1)\big) \subset (p_2)_\#\big(\pi_1(X_2, x_2)\big).$$

ning is wrong here

By the Map Lifting Lemma, we can lift p_1 to a (unique) map $h : X_1 \to X_2$ such that $h(x_1) = x_2$ and lift p_2 to a (unique) map $k : X_2 \to X_1$ such that $k(x_2) = x_1$. The maps k is the inverse of h, so that h is a homeomorphism, which proves that p_1 and p_2 are isomorphic.

EXAMPLE 6.4.7. This example, due to Zeeman, shows that for a covering space $p : X \to B$ with non locally path connected space X, the lift $\widetilde{f}$ of a map $f : A \to B$ (which always exists and is unique) may be discontinuous.

The spaces A, B, and X consist of a central circle, one (or two) half circles, one (or two) infinite sequences of segments with common end points as shown on the figure. Obviously all three of these spaces fail to be locally path connected at the points a, b, c, d, e.

FIGURE ??? Zeeman's example

The covering space $p : X \to B$ is obtained by wrapping the central circle of X around the central circle of B twice, and mapping the segments and arcs of X homeomorphically onto the corresponding segments and the arc of B. The map $f : A \to B$ (which also happens to be a covering) is defined exactly in the same way. Then we have

$$f_\#\big(\pi_1(A)\big) = p_\#\big(\pi_1(X)\big) \cong 2\mathbb{Z} \subset \mathbb{Z} \cong \pi_1(B),$$

but the map f has no continuous lift $\widetilde{f}$, because the condition $f \circ \widetilde{f} = p$ implies the uniqueness of $\widetilde{f}$, but this map is necessarily discontinuous at the points a and b.

EXERCISE 6.4.2. Can the identity map of $\mathbb{S}^1$ to $\mathbb{S}^1$ be lifted to the exponential covering $\mathbb{R} \to \mathbb{S}^1$?

EXERCISE 6.4.3. Prove that a map of $\mathbb{S}^1$ to $\mathbb{S}^1$ can be lifted to the exponential covering $\mathbb{R} \to \mathbb{S}^1$ if and only if its degree is zero.

EXERCISE 6.4.4. Describe the maps of $\mathbb{S}^1$ to $\mathbb{S}^1$ that can be lifted to the n-sheeted covering $w_n : \mathbb{S}^1 \to \mathbb{S}^1$.

EXERCISE 6.4.5. What maps of the circle to the torus $\mathbb{T}^2$ can be lifted to the universal covering $\mathbb{R}^2 \to \mathbb{T}^2$?

6.5. Coverings of surfaces and the Euler characteristic

In this section we compute the fundamental group of compact surfaces, use it to investigate the covering spaces of surfaces, and investigate the behavior of the Euler characteristic under the corresponding covering maps.

6.5.1. Behavior of the Euler characteristic. Here we will see that for a covering of a surface by another surface, the Euler characteristics of the surfaces are directly related to the number of sheets of the covering. In fact, we have the following theorem.

THEOREM 6.5.1. *For any n-sheeted covering $p : M \to N$ of the compact surface N by the compact surface M, the Euler characteristics of the surfaces are related by the formula*

$$\boxed{\chi(N) \;=\; n \cdot \chi(M)}$$

PROOF. We will prove the theorem for triangulated surfaces, which does not restrict generality by Theorem ??? Let $U_1, \dots, U_k$ be a covering of N by neighborhoods whose inverse images consist of n disjoint sets each of which is mapped homeomorphically onto the corresponding U_i. Take a triangulation of N and subdivide it barycentrically until each 2-simplex is contained in one of the U_i. Using the projection p, pull back this triangulation to M. Then the inverse image of each 2-simplex of N will consist of n 2-simplices of M and we will have

$$\begin{aligned}\chi(M) &= V_M - E_M + F_M = nV_M - nE_M + nF_M \\ &= n(V_M - E_M + F_M) = n\chi(N),\end{aligned}$$

where V, E, and F (with subscripts) stand for the number of vertices, edges, and faces of the corresponding surface. □

EXAMPLE 6.5.2. There is no covering of the surface $M_{(4)}$ of genus 4 by the surface $M_{(6)}$ of genus 6. Indeed, the Euler characteristic of $M_{(4)}$ is -6, while the Euler characteristic of $M_{(6)}$ is -10, and -6 does not divide -10.

EXERCISE 6.5.1. Can the surface of genus 7 cover that of genus 5?

EXERCISE 6.5.2. State and prove a theorem similar to the previous one for coverings of graphs.

6.5.2. Covering surfaces by surfaces. After looking at some examples, we will use the previous theory to find the genus of surfaces that can cover a surface of given genus.

EXAMPLE 6.5.3. The surface $M_{(9)}$ of genus 9 is the 4-fold covering of the surface of genus $M_{(3)}$. To see this, note that $M_{(9)}$ has a $\mathbb{Z}_4$ rotational symmetry (this is clear from the figure) which can be regarded as a normal action of $\mathbb{Z}_4$ on $M_{(9)}$. Obviously, the corresponding quotient space is $M_{(3)}$.

FIGURE ??? A four-sheeted covering of surfaces

EXERCISE 6.5.3. Generalizing the previous example, construct a d-fold covering $p : X \to B$ of of the orientable surface of genus k by the orientable surface of genus $d(k-1)+1$.

THEOREM 6.5.4. *A compact orientable surface M covers a compact orientable surface N if and only if the Euler characteristic of N divides that of M.*

PROOF. If $\chi(N)$ does not divide $\chi(M)$, then M cannot cover N by the theorem in the previous subsection, which proves the "only if" part of the theorem. To prove the "if" part, assume that $\chi(M) = d \cdot \chi(N)$. Since the Euler characteristic of a surface can be expressed via its genus as $\chi = 2 - 2g$, we have $d(2-g_N) = 2-2g_M$, whence $g_M = dg_N - d - 1$. Now the construction from the previous exercise (which is a straightforward generalization of the one described in the previous example) completes the proof of the theorem. $\square$

EXERCISE 6.5.4. Describe all possible coverings of nonorientable surfaces by nonorientable surfaces.

EXERCISE 6.5.5. Describe all possible coverings of orientable surfaces by nonorientable ones.

6.6. Branched coverings of surfaces

In this and the next section, we study the theory of branched coverings (also called ramified coverings) of two-dimensional manifolds (surfaces). This is a beautiful theory, originally coming from complex analysis, but which has drifted into topology and, recently, into mathematical physics, where it is used to study such fashionable topics as moduli spaces and Gromov–Witten theory.

6.6.1. Main definitions. Suppose that M^2 and N^2 are two-dimensional manifolds. Recall that a continuous map $p : M^2 \to N^2$ is said to be a *covering* (with *fiber* Γ, where Γ is a fixed discrete space) if for every point $x \in N^2$ there exists a neighborhood U and a homeomorphism $\varphi : p^{-1}(U) \to U \times \Gamma$ such that the restriction of p to $p^{-1}(U)$ coincides with $\pi \circ \varphi$, where the map $\pi : U \times \Gamma \to U$ is the projection on the first factor. Then M^2 is called the *covering manifold* or *covering space*, while N^2 is the *base manifold* or *base*. If the fiber Γ consists of n points, then the covering p is said to be *n-fold*.

A continuous map $p : M^2 \to N^2$ is said to be a *branched* (or *ramified*) *covering* if there exists a finite set of points $x_1, \dots, x_n \in N^2$ such that the set $p^{-1}(\{x_1, \dots, x_n\})$ is discrete and the restriction of the map p to the set $M^2 - p^{-1}(\{x_1, \dots, x_n\})$ is a covering. In other words, after we delete a finite set of points, we get a covering. The points $x_1, \dots, x_n \in N^2$ that must be deleted are called the *branch points* of p. The following obvious statement not only provides an example of a branched covering, but shows how branched coverings behave near branch points.

PROPOSITION 6.6.1. *Let $D^2 = \{z \in \mathbb{C} : |z| \leqslant 1\}$ and let $p : D^2 \to D^2$ be the map given by the formula $p(z) = z^m$. Then p is an m-fold branched covering with unique branch point $z = 0$.* $\square$

The example in the proposition for different m describes the structure of an arbitrary branched covering near its branch points. Indeed, it turns out that if p is an n-fold branched covering and U is a sufficiently small disk neighborhood of a branch point, then $p^{-1}(U)$ consists of one or several disks on which p has the same structure as the map in the proposition (in general, with different values of m). We shall not prove this fact in the general case, but in all the examples considered it will be easy to verify that this is indeed the case. If in a small neighborhood of a point x of the covering manifold the covering map is equivalent to the map $z \mapsto z^m$, we shall say that x has *branching index* m. The following proposition is obvious.

PROPOSITION 6.6.2. *For any n-fold branched covering, the sum of branching indices of all the preimages of any branch point is equal to n.*

Here is another, less trivial, example of a branched covering, which will allow us to construct a branched covering of the open disk (actually, the open set with boundary an ellipse) by the open annulus.

PROPOSITION 6.6.3. *Consider the map $f : \mathbb{C}\backslash\{0\} \to \mathbb{C}$ given by the formula $f(z) = 2(z + 1/z)$. This map is a 2-fold branched covering with branch points*

± 4. *The preimages of these points are the points* ± 1, *and the branching index of each is* 2.

PROOF. The equation $2(z+1/z)=c$ is quadratic. Its discriminant $c^2/4-4$ vanishes iff $c=\pm 4$. This value is assumed by the function f when $z=\pm 1$. $\square$

PROPOSITION 6.6.4. *Let* p *be the restriction of the map* f *from the previous proposition to the annulus* $C=\{z\in\mathbb{C}:1/2<|z|<2\}$. *If* $z=\rho e^{i\varphi}$, *then we have*

$$p(z)=2((\rho+1/\rho)\cos\varphi+i(\rho-1/\rho)\sin\varphi),$$

so that the image of the annulus C *is the set of points located inside the ellipse* (*see Fig. ??*)*:* $\{\rho=5\cos\varphi+3i\sin\varphi,\ 0\leqslant\varphi<2\pi\}$. $\square$

FIGURE ?? Branched covering of the ellipse by the annulus

A more geometric description of p is the following. Imagine the open annulus C as a sphere with two holes (closed disks) and an axis of symmetry l (Fig.??). Let us identify points of the set C symmetric with respect to l. It is easy to see that the resulting space is homeomorphic to the open disk D^2. The quotient map $p:C\to D^2$ thus constructed is a 2-fold branched covering with two branch points (the intersection points of l with the sphere).

FIGURE ?? Sphere with two holes and symmetry axis

EXERCISE 6.6.1. Prove that if the base manifold M^2 of a branched covering $p:N^2\to M^2$ is orientable, then so is the covering manifold N^2.

THEOREM 6.6.5. *Let* M_g^2 *be the sphere with* g *handles. Then there exists a branched covering* $p:M_g^2\to S^2$.

First proof. Consider a copy of the sphere with g handles with an axis of symmetry l (Fig.??). Identify all pairs of points symmetric with respect to l. The resulting quotient space is (homeomorphic to) the ordinary sphere $\mathbb{S}^2$. The natural projection $p:M_g^2\to\mathbb{S}^2$ is a 2-fold branched covering which has $2g+2$ branch points. $\square$

FIGURE ?? Branched covering of the 2-sphere

Second proof. Consider a *triangulation* of the manifold M_g^2. (This means that M_g^2 is cut up into (curvilinear) triangles, any two of which either intersect along a common side, or intersect in a common vertex, or have no common points.) Let $A_1, \ldots, A_n$ be the vertices of the triangulation. On the sphere $\mathbb{S}^2$ choose n points $B_1, \ldots, B_n$ situated in general position in the following sense: no three of them lie on one and the same great circle and no two are antipodes. Then any three points B_i, B_j, B_k uniquely determine a spherical triangle Δ_1. Suppose Δ_2 is the closure of its complement $S^2 - \Delta_1$; then Δ_2 is also homeomorphic to the triangle. Therefore there exist homeomorphisms

$$f_1 : A_iA_jA_k \to \Delta_1, \qquad f_2 : A_iA_jA_k \to \Delta_2$$

that are linear in the following sense. We can assume that the length of a curve is defined both on the manifold M_g^2 and on the sphere $\mathbb{S}^2$; we require that an arbitrary point X divide the arc A_pA_q in the same ratio as the point $f_r(X), r = 1, 2$, divides the arc B_pB_q.

Let us fix orientations of M_g^2 and S^2. The orientations of the triangles $A_iA_jA_k$ and $B_iB_jB_k$ induced by their vertex order may agree with or be opposite to that of M_g^2 and $\mathbb{S}^2$. If both orientations agree, or both are opposite, then we map $A_iA_jA_k$ onto $\Delta_1 = B_iB_jB_k$ by the homeomorphism f_1. If one orientation agrees and the other doesn't, we map $A_iA_jA_k$ onto Δ_2 (the complement to $B_iB_jB_k$) via f_2. Defining such maps on all the triangles of the triangulation of M_g^2, we obtain a map $f : M_g^2 \to \mathbb{S}^2$. We claim that this map is a branched covering.

For each interior point x_0 of triangle $A_iA_jA_k$ there obviously exists a neighborhood $U(x_0)$ mapped homeomorphically onto its image. Let us prove that this is not only the case for interior points of the triangles, but also for inner points of their sides. Indeed, let the side A_iA_j belong to the two triangles $A_iA_jA_k$ and $A_iA_jA_l$. On the sphere $\mathbb{S}^2$, the great circle passing through the points B_i and B_j may either separate the points B_k and B_l or not separate them. In the first case we must have used the maps f_1 and f_1 (or the maps f_2 and f_2), in the second one f_1 and f_2 (or f_2 and f_1). In all cases a sufficiently small neighborhood of a point chosen inside A_iA_j will be mapped bijectively (and hence homeomorphically) onto its own image (Fig.??). So at all points except possibly the vertices $A_1, \ldots, A_n$ we have a covering, so f is a branched covering. $\Box$

FIGURE ?? Structure of f at inner points of the sides

In fact not all the points $B_1, \dots, B_n$ are necessarily branch points, although some must be when $g > 0$. How few can there be? The answer is contained in the next statement, and surprisingly does not depend on g.

THEOREM 6.6.6. *Let M_g^2 be the sphere with g handles, where $g \geqslant 1$. Then there exists a branched covering $p : M_g^2 \to \mathbb{S}^2$ with exactly three branch points.*

PROOF. Choose an arbitrary triangulation of the manifold M_g^2 and take its baricentric subdivision, i.e., subdivide each triangle into 6 triangles by its three medians. To the vertices of the baricentric subdivision assign the numbers $0, 1, 2$ as follows:

0 to the vertices of the initial triangulation;
1 to the midpoints of the sides;
2 to the baricenters of the triangles.

If the orientation of some triangle 012 induced by its vertex order agrees with that of the manifold M_g^2, then we paint the triangle black, otherwise we leave it white (Fig.??).

FIGURE ?? Black and white coloring of the baricentric subdivision

We can regard the sphere $\mathbb{S}^2$ as the union of two triangles glued along their sides; denote their vertices by $0', 1', 2'$ and paint one of these triangles black, leaving the other white. Now we map each white triangle 012 from the baricentric subdivision in M_g^2 to the white triangle $0'1'2'$ in $\mathbb{S}^2$, and each black one to the black $0'1'2'$. In more detail this map was described in the second proof of Theorem ??, where it was established that this map is a branched covering. The three branch points are of course $0', 1', 2'$. $\square$

6.7. Riemann–Hurwitz formula

The formula that we prove in this section relates the topological properties of the base and covering manifolds in a branched covering with the branching indices. The main tool involved is the Euler characteristic, and we begin by recalling some of its properties.

6.7.1. Some properties of the Euler characteristic for surfaces. The properties that we need were proved in the previous chapter, but we restate them here for completeness as exercises.

Recall that a triangulation K' is said to be a *subdivision* of the triangulation K if any simplex of K' is the union of simplices from K. Two triangulations K_1 and K_2 of a two-dimensional manifold are called *transversal* if their edges intersect transversally at a finite number of points. Any two triangulations of a 2-manifold can be made transversal by a small move (see ??)

EXERCISE 6.7.1. (a) Verify that the Euler characteristic of a 2-manifold does not change when we pass to a subdivision of its triangulation.

(b) Prove that any two transversal triangulations of a 2-manifold have a common subdivision. Using (a) and (b), show that the Euler characteristic for 2-manifolds does not depend on the choice of the triangulation.

EXERCISE 6.7.2. Suppose a surface M is cut into two components A and B by a circle. Prove that in this case $\chi(M) = \chi(A) + \chi(B)$.

EXERCISE 6.7.3. Show that if the surface M can be obtained from the surface F by adding a handle (see Fig.??), then $\chi(M) = \chi(F) - 2$

FIGURE 21.1 Adding a handle

EXERCISE 6.7.4. Prove that if M_g^2 is the sphere with g handles, then $\chi(M_g^2) = 2 - 2g$.

Thus the topological type of any oriented compact 2-manifold without boundary is entirely determined by its Euler characteristic.

6.7.2. Statement of the Riemann-Hurwitz theorem. The main result of this section is the following

THEOREM 6.7.1 (Riemann-Hurwitz formula). *Suppose $p : M^2 \to N^2$ is an n-fold branched covering of compact 2-manifolds, $y_1, \dots, y_l$ are the preimages of the branch points, and $d_1, \dots, d_l$ are the corresponding branching indices. Then*

$$\boxed{\chi(M^2) + \sum_{i=1}^{l} (d_i - 1) = n\chi(N^2).}$$

The proof of Theorem ?? will be easier to understand if we begin by recalling its proof in the particular case of non-branched coverings (see Theorem ?? in the previous section) and work out the following exercise.

EXERCISE 6.7.5. a) Let $p : M_g^2 \to N_h^2$ be the covering of the sphere with h handles by the sphere with g handles. Prove that $g - 1$ is divisible by $h - 1$.

b) Suppose that $g, h \geqslant 2$ and $g - 1$ is divisible by $h - 1$. Prove that there exists a covering $p : M_g^2 \to N_h^2$.

6.7.3. Proof of the Riemann–Hurwitz formula. First let us rewrite formula ?? in a more convenient form. After an appropriate renumbering of the branching indices, can assume that $d_1, \dots, d_{a_1}$ are the branching indices of all the preimages of one branch point, $d_{a_1+1}, \dots, d_{a_1+a_2}$ are the branching indices of all the preimages of another branch point, and so on. Since (see ??)

$$d_1 + \cdots + d_{a_1} = d_{a_1+1} + \cdots + d_{a_1+a_2} = \cdots = n,$$

we obtain

$$\sum_{i=1}^{l}(d_i - 1) = (n - a_1) + (n - a_2) + \cdots = kn - a_1 - a_2 - \cdots - a_k,$$

where k is the number of branch points and a_i is the number of preimage points of the ith branch point. Hence formula ?? can be rewritten as

$$\chi(M^2) = n(\chi(N^2) - k) + a_1 + \cdots + a_k.$$

It will be more convenient for us to prove the Riemann–Hurwitz formula in this form.

The manifolds M^2 and N^2 may be presented as follows

$$M^2 = A_M \cup B_M \quad \text{and} \quad N^2 = A_N \cup B_N,$$

where A_N is the union of the closures of small disk neighborhoods of all the branch points, A_M is the inverse image of A_N, while B_M and B_N are the closures of the complements $M^2 - A_M$ and $N^2 - A_N$. The sets $A_M \cap B_M$ and $A_N \cap B_N$ consist of nonintersecting circles, and so we can use formula from Exercise ??. As the result we get

$$\chi(M^2) = \chi(A_M) + \chi(B_M), \quad \chi(N^2) = \chi(A_N) + \chi(B_N).$$

The restriction of the map p to the set B_M is a (nonbranched) covering, so by Theorem ?? we have

$$\chi(B_M) = n\chi(B_N).$$

The set A_N consists of k disks, while A_M consists of $a_1 + \cdots + a_k$ disks. Therefore

$$\chi(A_M) = a_1 + \cdots + a_k, \quad \chi(A_N) = k.$$

Combining the displayed formulas, we get

$$\chi(M^2) = a_1 + \cdots + a_k + n\chi(B_N) = a_1 + \cdots + a_k + n(\chi(N^2) - k),$$

which is the required formula ??. This completes the proof of the Riemann–Hurwitz formula. □

6.7.4. Some applications. The Riemann–Hurwitz formula has numerous applications. The first one that we discuss has to do with Theorem ??, which asserts the existence of a branched covering $p : M_g^2 \to \mathbb{S}^2$ with exactly 3 branch points (when $g \geqslant 1$). Can this number be decreased? The answer is ‘no’, as the next statement shows.

THEOREM 6.7.2. *If $g \geqslant 1$, there exists no branched covering $p : M_g^2 \to \mathbb{S}^2$ of the sphere by the sphere with g handles having less than 3 branch points.*

PROOF. By formula ??, we have

$$2-2g=\chi(M_g^2)=n(\chi(S^2)-k)+a_1+\cdots+a_k=n(2-k)+a_1+\cdots+a_k.$$

If $k \leqslant 2$, then $n(2-k) \geqslant 0$ and hence $n(2-k)+a_1+\cdots+a_k>0$, because the relations $n(2-k)=0$ and $a_1+\cdots+a_k=0$, i.e., $k=0$, cannot hold simultaneously. Thus $2-2g>0$, hence $g=0$, contradicting the assumption of the theorem. Therefore $k>2$ as asserted. □

EXERCISE 6.7.6. Prove that if $p:\mathbb{D}^2\to\mathbb{D}^2$ is a branched covering of the disk by the disk with exactly one branch point, then the preimage of the branch point consists of one point.

6.7.5. Genus of complex algebraic curves. The Riemann–Hurwitz formula can be applied to the computation of the genus of algebraic curves in $\mathbb{C}P^2$. We begin with the necessary background material. An *algebraic curve* of degree n in $\mathbb{C}P^2$ is the set of points satisfying the homogeneous equation

$$F(x,y,z)=\sum_{i+j+k=n} a_{ij}x^iy^jz^{n-i-j}, \quad x,y,z\in\mathbb{C}P^2.$$

When $z=1$ and $x,y\in\mathbb{R}$, we get a *plane algebraic curve*:

$$\sum_{i+j\leqslant n} a_{ij}x^iy^j=0.$$

If the gradient

$$\operatorname{grad}F=\left(\frac{\partial F}{\partial x},\frac{\partial F}{\partial y},\frac{\partial F}{\partial z}\right)$$

does not vanish at all points of an algebraic curve in $\mathbb{C}P^2$, then the curve is called *nonsingular*. If the polynomial F is irreducible, i.e., cannot be represented as the product of two homogeneous polynomials of lesser degree, then the curve is said to be *irreducible*. It can be proved that any nonsingular irreducible algebraic curve in $\mathbb{C}P^2$ is homeomorphic to the sphere with g handles for some $g\geqslant 0$; the nonnegative integer g is called the *genus* of the curve.

PROPOSITION 6.7.3. *Fermat's curve $x^n+y^n+z^n=0$ is of genus $(n-1)(n-2)/2$.*

PROOF. First note that Fermat's curve $\Gamma\subset\mathbb{C}P^2$ is nonsingular, because

$$\operatorname{grad}F=n\left(x^{n-1},y^{n-1},z^{n-1}\right)\neq 0 \quad \forall(x:y:z)\in\mathbb{C}P^2.$$

Now consider the map $p:\mathbb{C}P^2-\{(0:0:1)\}\to\mathbb{C}P^1$ that takes the point $(x:y:z)\in\mathbb{C}P^2$ to $(x:y)\in\mathbb{C}P^1$. The point $(0:0:1)$ does not lie on the curve Γ, therefore the map p induces the map $p':\Gamma\to\mathbb{C}P^1=\mathbb{S}^2$. The inverse image of the point $(x_0:y_0)\in\mathbb{C}P^1$ consists of all points $(x_0:y_0:z)\in\mathbb{C}P^2$ such that $z^n=-(x_0^n+y_0^n)$. When $x_0^n+y_0^n\neq 0$, the inverse image consists of n points, when $x_0^n + y_0^n = 0$, of only one. Hence p' is an n-fold branched covering with branch points $(1 : \varepsilon_n) \in \mathbb{C}P^1$, where ε_n is a root of unity of degree n. So there are n branch points, and the inverse image of each consists of one point. According to formula ??, we get

$$\chi(\Gamma) = n(\chi(S^2) - n) + n = n(2 - n) + n = -n^2 + 3n.$$

Therefore,

$$g = \frac{2 - \chi(\Gamma)}{2} = \frac{n^2 - 3n + 2}{2} = \frac{(n-1)(n-2)}{2}.$$

□

PROPOSITION 6.7.4. *The hyperelliptic curve $y^2 = P_n(x)$, where P_n is a polynomial of degree $n \geqslant 5$ without multiple roots, is of genus*

$$\left[\frac{n+1}{2}\right] - 1.$$

REMARK 6.7.5. This statement is also true for $n < 5$. When $n = 3, 4$ the curve $y^2 = P_n(x)$ is called *elliptic*.

PROOF. Let $P_n(x) = a_0 + \cdots + a_n x^n$. Then the hyperelliptic curve Γ in $\mathbb{C}P^2$ is given by the equation

$$y^2 z^{n-2} = \sum_{k=0}^{n} a_k x^k z^{n-k}.$$

For the curve Γ, we have $\operatorname{grad} F = 0$ at the point $(0 : 1 : 0) \in \Gamma$, so the hyperelliptic curve is singular.

Consider the map

$$p : \mathbb{C}P^2 - \{(0:1:0)\} \to \mathbb{C}P^1, \quad \mathbb{C}P^2 \ni (x : y : z) \mapsto (x : z) \in \mathbb{C}P^1.$$

Let $p' : \Gamma - \{(0:1:0)\} \to \mathbb{C}P^1$ be the restriction of p. We claim that for $x = 1$ the preimage of the point $(x : z)$ tends to the singular point $(0 : 1 : 0) \in \Gamma$ as $z \to 0$. Indeed, by ?? we have $y^2 \approx a_n z^{2-n} \to \infty$, so that

$$(1 : y : z) = (1/y : 1 : z/y) \to (0 : 1 : 0).$$

Therefore the map p' can be extended to a map of the whole curve, $p' : \Gamma \to \mathbb{C}P^1$, by putting $p'((0:1:0)) = (1:0)$.

In order to find the inverse image under p' of the point $(x_0 : z_0) \in \mathbb{C}P^1$ when $z_0 \neq 0$, we must solve the equation

$$y^2 = z_0^2 \sum_k a_k \cdot \left(\frac{x_0}{z_0}\right)^k = z_0^2 \, P_n\left(\frac{x_0}{z_0}\right).$$

If x_0/z_0 is not a root of the polynomial P_n, then this equation has exactly two roots. Therefore the map $p' : \Gamma \to \mathbb{C}P^1$ is a double branched covering, the branch points being $(x_0 : z_0) \in \mathbb{C}P^1$, where x_0/z_0 is a root of P_n and, possibly, the point

$(1:0)$. We claim that $(1:0)$ is a branch point iff n is odd. To prove this claim, note that for small z the preimage of the point $(1:z)$ consists of points of the form $(1:y:z)$, where $y^2 \approx a_n z^{2-n}$. Let $z = \rho e^{i\phi}$. When ϕ varies from 0 to 2π, i.e., when we go around the point $(1:0) \in \mathbb{C}P^1$, the argument of the point $y \in \mathbb{C}$ changes by

$$(2-n)2\pi/2 = (2-n)\pi.$$

Therefore for odd n the number y changes sign, i.e., we switch to a different branch, while for n even y does not change, i.e., we stay on the same branch.

Thus the number of branch points is $2[(n+1)/2]$. Let g be the genus of the curve Γ. Then, according to ??,

$$2-2g = 2\left(2-2\left[\frac{n+1}{2}\right]\right) + 2\left[\frac{n+1}{2}\right], \quad \text{i.e.,} \quad g = \left[\frac{n+1}{2}\right] - 1.$$

□

6.8. Problems

EXERCISE 6.8.1. Give an example of the covering of the wedge sum of two circles which is not normal.

EXERCISE 6.8.2. Prove that any nonorientable surface possesses a double covering by an oriented one.

EXERCISE 6.8.3. Prove that any subgroup of a free group is free by using covering spaces.

EXERCISE 6.8.4. Prove that if G is a subgroup of a free group F of index $k := [F:G] \leq \infty$, then its rank is given by $\operatorname{rk} G = k(\operatorname{rk} F - 1) + 1$

REMARK 6.8.1. Note that the purely group-theoretic proofs of the two theorems appearing in the previous two exercises, especially the first one, are quite difficult and were hailed, in their time, as outstanding achievements. As the reader should have discovered, their topological proofs are almost trivial.

EXERCISE 6.8.5. Prove that the free group of rank 2 contains (free) subgroups of any rank n (including $n = \infty$)

EXERCISE 6.8.6. Give two examples of nonisomorphic three-sheeted coverings of the wedge of two circles,

EXERCISE 6.8.7. Suppose the graph G'' covers the graph G'. What can be said about their Euler characteristics?

EXERCISE 6.8.8. Prove that any double (i.e., two-sheeted) covering is normal. To what group-theoretic statement does this fact correspond?

EXERCISE 6.8.9. Prove that any three-sheeted covering of the sphere with two handles cannot be normal (i.e., regular in the traditional terminology).

EXERCISE 6.8.10. Construct a double covering of n-dimensional projective space $\mathbb{R}P^n$ by the sphere and use it to prove that
(i) $\mathbb{R}P^n$ is orientable for even n and orientable for odd n.
(ii) $\pi_1(\mathbb{R}P^n) = \mathbb{Z}_2$ and $\pi_k(\mathbb{R}P^n) \cong \pi_k(\mathbb{S}^n)$ for $n \geq 2$.

EXERCISE 6.8.11. Prove that all the coverings of the torus are normal (i.e., regular in the traditional terminology) and describe them.

EXERCISE 6.8.12. Find the universal covering of the Klein bottle K and use it to compute the homotopy groups of K.

EXERCISE 6.8.13. Can the torus double cover the Klein bottle?

EXERCISE 6.8.14. Prove that any deck transformation of an arbitrary (not necessarily normal) covering is entirely determined by one point and its image.

MATH 527: GEOMETRY/TOPOLOGY I.

FALL 2006

A.Katok

HOMEWORK #1; September 11, 2006

Topological Spaces

Due on Monday September 18

1. Write complete proof of Proposition 1.1.13: For any set X and any collection $\mathcal{C}$ of subsets of X there exists a unique weakest topology for which all sets from $\mathcal{C}$ are open.

2.(Ex. 1.1.1.) How many different (non-homeomorphic) topologies are there on the 2–element set and on the 3–element set?

3.(Ex. 1.3.2.)Prove that the sphere $\mathbb{S}^2$ with two points removed is homeomorphic to the infinite cylinder $C := \{(x, y, z) \in \mathbb{R}^3 | x^2+y^2 = 1\}$.

4.(Example 1.3.3.+ Ex. 1.3.4 & 1.3.5.) Prove that the following three constructions of the n-torus $\mathbb{T}^n$ produce homeomorphic topological spaces:

- Product of n copies of the circle
- The following subset of $\mathbb{R}^{2n}$:

  $$\{(x_1, \ldots x_{2n}) : x_{2i-1}^2 + x_{2i}^2 = 1,\ i = 1, \ldots, n.\}$$

  with the induced topology.
- The identification space of the unit n–cube I^n:

  $$\{(x_1, \ldots, x_n) \in \mathbb{R}^n : 0 \leq x_i \leq 1,\ i = 1, \ldots n\}.$$

  where any two points are identified if all of their coordinates but one are equal and the remaining one is 0 for one point and 1 for another.

5.(Ex. 1.10.4 & 1.10.5.) Consider the *profinite* topology on $\mathbb{Z}$ in which open sets are defined as unions (not necessarily finite) of non-constant arithmetic progressions

a) Prove that this defines a topology.

b) Let $\mathbb{T}^\infty$ be the product of countably many copies of the circle with the product topology. Define the map $\varphi : \mathbb{Z} \to \mathbb{T}^\infty$ by

$$\varphi(n) = (\exp(2\pi in/2), \exp(2\pi in/3), \exp(2\pi in/4), \exp(2\pi in/5), \ldots)$$

Show that the map φ is injective and that the pullback topology on $\varphi(\mathbb{Z})$ coincides with its profinite topology.

MATH 527: GEOMETRY/TOPOLOGY I

FALL 2006

A.Katok

HOMEWORK #2; September 18, 2006

Factor–spaces, Separation, Compactness

Due on Monday September 25

6. Ex 1.3.9.

7. Ex 1.3.11.

8. Ex 1.4.3.

9. Ex 1.10.6.

10. Ex 1.5.3.

“‘Extra credit” problems

You may submit solutions until October 6.

E1. Ex 1.10.9.

E2. Ex 1.10.10.

MATH 527: GEOMETRY/TOPOLOGY I

FALL 2006

A.Katok

HOMEWORK # 3; September 25, 2006

Compactness, Connectedness, Manifolds

Due on Monday, October 2

ATTENTION: NUMBERS FROM REVISED VERSION of "Chapter 1"

11. Ex 1.5.5
Hint: You may use induction in n.

12. Ex. 1.10.7

13. Ex 1.7.2.

14. Ex 1.10.14.

15. Ex 1.8.6.

"'Extra credit" problem

You may submit solutions until October 13.

E3. Ex 1.8.3.

MATH 527: GEOMETRY/TOPOLOGY I

FALL 2006

A.Katok

HOMEWORK # 4; October 2, 2006

Manifolds, graphs, beginning homotopy

Due on Monday October 9

16. Consider the torus $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$ and let S be the quotient space obtained by identifying orbits of the map $I : x \mapsto -x$. Prove that S is homeomorphic to the sphere $\mathbb{S}^2$.

17. Ex 2.3.1.

18. Ex. 2.12.1.

19. Ex 2.12.2.

20. Ex 2.12.4.

"‘Extra credit” problems

You may submit solutions until October 27.

E4. Consider regular $2n$-gon and identify pairs of opposite side by the corresponding parallel translations. Prove that the identification space is a topological manifold. Prove that the manifolds obtained by this construction from the $4n$-gon and and $4n+2$-gon are homeomorphic.

E5. Prove that the manifold of the previous exercise is homeomorphic to the surface of the sphere to which n "handles" are attached, or, equivalently, to the surface of n tori joint into a "chain" (Figure 1.8.1 illustrates this for $n = 1$ and $n = 3$.

MATH 527: GEOMETRY/TOPOLOGY I

FALL 2006

A.Katok

HOMEWORK # 5: October 9, 2006

Fundamental group, covering spaces

Due on Monday October 16

21. Prove that for any path connected topological space X we have $\pi_1(\text{Cone(X)}) = 0$.

22. Consider the following map f of the torus $\mathbb{T}^2$ into itself:

$$f(x, y) = (x + \sin 2\pi y, 2y + x + 2\cos 2\pi x) \pmod 1.$$

Describe the induced homomorphism f_* of the fundamental group.

23. Let $X = \mathbb{R}^2 \setminus \mathbb{Q}^2$. Prove that $\pi_1(X)$ is uncountable.

24. Let X be the quotient space of the disjoint union of $\mathbb{S}^1$ and $\mathbb{S}^2$ with a pair of points $x \in S^1$ and $y \in S^2$ identified. Calculate $\pi_1(X)$.

25. Describe two-fold coverings of

(1) the (open) Möbius strip by the open cylinder $\mathbb{S}^1 \times \mathbb{R}$;
(2) the Klein bottle by the torus $\mathbb{T}^2$.

26. Prove that the real projective space $\mathbb{R}P(n)$ is not simply connected.

Hint: Use the fact that $\mathbb{R}P(n)$ is the sphere $\mathbb{S}^n$ with diametrically opposed points identified.

MATH 527: GEOMETRY/TOPOLOGY I

FALL 2006

A.Katok

HOMEWORK # 6; October 23 2006

Hopf fibration, Cantor sets, differentiable manifolds

Due on Monday October 30

26. For any finite cyclic group C there exists a compact connected three-dimensional manifold whose fundamental group is isomorphic to C.

Hint: Use the Hopf fibration.

27. For any abelian finitely generated group A there exists a compact manifold whose fundamental group is isomorphic to A.

Hint: Use the fact that any finitely generated abelian group is the direct product of cyclic groups (finite and infinite).

28. Prove that complex projective space $\mathbb{C}P(n)$ is simply connected for every n.

29. Introduce a metric d on the Cantor set C (generating the Cantor set topology) such that (C, d) cannot be isometrically embedded to $\mathbb{R}^n$ for any n.

30. Construct a smooth atlas of the projective space $\mathbb{R}P(3)$ with as few charts as possible.

"'Extra credit" problems

You may submit solutions until November 14.

E6. Introduce a metric d on the Cantor set C such that (C, d) is not Lipschitz equivalent to a subset of $\mathbb{R}^n$ for any n.

E7. Prove that for any finite graph G, $\pi_n(G) = 0$ for any $n \geq 2$.

MATH 527: GEOMETRY/TOPOLOGY I

FALL 2006

A.Katok

HOMEWORK # 7; October 30, 2006

Differentiable manifolds; diffeomorphisms, submanifolds

Due on Monday, November 6

31. Construct an explicit diffeomorphism between $\mathbb{R}^n$ and the open unit ball B^n.

32. Prove that any convex open set in $\mathbb{R}^n$ is diffeomorphic to $\mathbb{R}^n$.

33. Prove that the following three smooth structures on the torus $\mathbb{T}^2$ are equivalent, i.e. the torus provided with any of these structure is diffeomorphic to the one provided with another:

- $\mathbb{T}^2 = \mathbb{S}^1 \times \mathbb{S}^1$ with the product structure;
- $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$ with the factor-structure;
- The embedded torus of revolution in $\mathbb{R}^3$
$$\mathbb{T}^2 = \left\{(x, y, z) \in \mathbb{R}^3 : \left(\sqrt{x^2 + y^2} - 2\right)^2 + z^2 = 1\right\}$$
with the submanifold structure.

34. Prove that the n-dimensional torus in $\mathbb{R}^{2n}$:
$$x_{2k-1}^2 + x_{2k}^2 = \frac{1}{n}, \quad k = 1, \dots, n$$
is a smooth submanifold of the $(2n-1)$-dimensional sphere
$$\sum_{i=1}^{2n} x_i^2 = 1.$$

35. Prove that the upper half of the cone
$$x^2 + y^2 = z^2, \quad z \geq 0$$
is not a submanifold of $\mathbb{R}^3$, while the punctured one
$$x^2 + y^2 = z^2, \quad z > 0$$
is a submanifold of $\mathbb{R}^3$.

MATH 527: GEOMETRY/TOPOLOGY I

FALL 2006

A.Katok

HOMEWORK # 8; November 10, 2006

Tangent bundles, examples of Lie groups, orientability

Due on Monday November 20

36. Prove that the tangent bundle to the three-dimensional sphere $\mathbb{S}^3$ is diffeomorphic to the direct product $\mathbb{S}^3 \times \mathbb{R}^3$.

37. Find a natural smooth group structure on the sphere $\mathbb{S}^3$.

38. Prove that real projective spaces $\mathbb{R}P(n)$ are orientable for odd n and non-orientable for even n.

39. Prove that complex projective spaces $\mathbb{C}P(n)$ are orientable.

40. Prove that there exists a non-vanishing smooth vector field on any odd–dimensional sphere $\mathbb{S}^{2n-1}$.

41. Prove that the group $SL(2, \mathbb{R})$ of 2×2 matrices with determinant one is homotopy equivalent to the circle.

MATH 527: GEOMETRY/TOPOLOGY I

FALL 2006

A.Katok

HOMEWORK # 9; November 20, 2006

Complex manifolds, Lle groups

Due on Monday November 27

42. Give a detailed proof that any complex manifold is orientable.

43. Find a polynomial in two complex variables whose zero set is a complex curve homeomorphic to the sphere with two handles.

44. Let M is a complex manifold and suppose X is a nonvanishing vector field on M. Prove that there exists another nonvanishing vector field Y linearly independent of X.

45. Represent the torus $\mathbb{T}^n$ as a linear group.

46. Prove that the group of affine transformations of $\mathbb{R}^n$ is isomorphic to a Lie subgroup of $GL(n+1, \mathbb{R})$. Calculate its dimension.

MATH 527: GEOMETRY/TOPOLOGY I

FALL 2006

A.Katok

HOMEWORK # 10

Surfaces I: triangulations, functions, Möbius caps

Due on Monday December 4

47. Prove that there exists a triangulation of the projective plane with any given number $N > 4$ of vertices.

48. Prove that minimal number of vertices in a triangulation of the torus is seven.

Hint: Use Euler theorem.

49. Prove that for any triangulation $\mathcal{T}$ of a surface there exists a smooth function whose local maxima are vertices of $\mathcal{T}$ and which has exaclty one saddle on each edge of $\mathcal{T}$, exaclty one local minimum inside each face of $\mathcal{T}$ and no more critical points.

50. Construct a smooth function of the torus with three critical points.

In the next two problems you must describe the homeomorphisms explicitly and not refer to the general theorem about classification of surfaces.

51. Given a surface M attaching a *Möbius cap* consists of deleting a small disk and identifying the resulting boundary circle with the boundary of a Möbius strip.

Prove that the sphere with two Möbius caps attached is homeomorphic to the Klein bottle.

52. Prove that sphere with three Möbius caps attached is homeomorphic to the torus with a Möbius cap attached.

MATH 527: GEOMETRY/TOPOLOGY I

FALL 2006

A.Katok

HOMEWORK # 11

Surfaces II: Classification, fundamental group, covering spaces

Due on Monday December 11

You cannot use classification of surfaces for the next two problems

53. Prove that attaching a handle decreases Euler characteristic of a compact surface (with or without boundary) by two.

54. Prove that attaching a Möbius cap decreases Euler characteristic of a compact surface (with or without boundary) by one.

55. Prove that the only compact surfaces from the standard list (spheres with handles, Möbius caps, and holes) which have abelian fundamental group are the sphere, the closed cylinder, the closed M öbius strip, the projective plane, and the torus.

57. Prove that any compact covering space for the torus is another torus.

58. Prove that the orientable surface of genus m is a covering space for an orientable surface of genus $n \geq 2$ if and only if $m > n$ and $n - 1$ divides $m - 1$.